Fundação Getulio Vargas

School of Applied Mathematics

Advanced Monte Carlo Methods

Pierre Del Moral

Introduction

Monte Carlo integration

These lectures are concerned with advanced Monte Carlo methods, including mean field type particle integration theory.

In the last three decades, this topic has become one of the most active contact points between pure and applied probability theory, Bayesian inference, statistical machine learning, information theory, theoretical chemistry and quantum physics, financial mathematics, signal processing, risk analysis, and several other domains in engineering and computer sciences.

The origins of Monte Carlo simulation certainly start with the seminal paper of N. Metropolis and S. Ulam in the late 1940s [452]. Inspired by the interest of his colleague S.Ulam in the poker game N. Metropolis, coined, the term "Monte Carlo Method" in reference to the "capital" of Monaco well known as the European city for gambling.

The first systematic use of Monte Carlo integration was developed by these physicists in the Manhattan Project of Los Alamos National Laboratory, to compute ground state energies of Schödinger's operators arising in thermonuclear ignition models. It is also not surprising that the development of these methods goes back to these early days of computers. For a more thorough discussion on the beginnings of the Monte Carlo method, we refer to the article by N. Metropolis [453].

As its name indicates, Monte Carlo simulation is, in the first place, one of the largest, and most important, numerical techniques for the computer simulation of mathematical models with random ingredients. Nowadays, these simulation methods are of current use in computational physics, physical chemistry, and computational biology for simulating the complex behavior of systems in high dimension. To name a few, there are turbulent fluid models, disordered physical systems, quantum models, biological processes, population dynamics, and more recently financial stock market exchanges. In engineering sciences, they are also used to simulate the complex behaviors of telecommunication networks, queuing processes, speech, audio, or image signals, as well as radar and sonar sensors.

Note that in this context, the randomness reflects different sources of model uncertainties, including unknown initial conditions, misspecified kinetic parameters, as well as the external random effects on the system. The repeated random samples of the complex system are used for estimating some *averaging* type property of some phenomenon.

Monte Carlo integration theory, including Markov chain Monte Carlo algorithms (abbreviated MCMC), sequential Monte Carlo algorithms (abbreviated SMC), and mean field interacting particle methods (abbreviated IPS) are also used to sample complex probability distributions arising in numerical probability, and in Bayesian statistics. In this context, the random samples are used for computing deterministic multidimensional *integrals*.

In other situations, stochastic algorithms are also used for solving complex *estimation problems*, including inverse type problems, global optimization models, posterior distributions calculations, nonlinear estimation problems, as well as statistical learning questions (see for instance [80, 99, 172, 288, 542]). We underline that in this situation, the randomness depends on the design of the stochastic integration algorithm, or the random search algorithm.

In the last three decades, these extremely flexible Monte Carlo algorithms have been developed

in various forms mainly in applied probability, Bayesian statistics, and in computational physics. Without any doubt, the most famous MCMC algorithm is the Metropolis-Hastings model presented in the mid-1960s by N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller in their seminal article [454].

This rather elementary stochastic technique consists in designing a reversible Markov chain, with a prescribed target invariant measure, based on sequential acceptance-rejection moves. Besides its simplicity, this stochastic technique has been used with success in a variety of application domains. The Metropolis-Hastings model is cited in *Computing in Science and Engineering* as being in the top 10 algorithms having the "greatest influence on the development and practice of science and engineering in the 20th century."

As explained by N. Metropolis and S. Ulam in the introduction of their pioneering article [452], the Monte Carlo method is, "essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences."

In this connection, we emphasize that any evolution model in the space of probability measures can always be interpreted as the distributions of random states of Markov processes. This key observation is rather well known for conventional Markov processes and related linear evolution models.

More interestingly, *nonlinear* evolution models in distribution spaces can also be seen as the laws of Markov processes, but their evolution interacts in a nonlinear way with the distribution of the random states. The random states of these Markov chains are governed by a flow of complex probability distributions with often unknown analytic solutions, or sometimes too complicated to compute in a reasonable time. In this context, Monte Carlo and mean field methods offer a catalog of rather simple and cheap tools to simulate and to analyze the behavior of these complex systems.

These two observations are the stepping stones of the mean field particle theory developed in these lectures.

Mean field simulation

The theory of mean field interacting particle models had certainly started by the mid-1960s, with the work of H.P. McKean Jr. on Markov interpretations of a class of nonlinear parabolic partial differential equations arising in fluid mechanics [445]. We quote an article by T.E. Harris and H. Kahn [338], published in 1951, using mean field type and heuristic-like splitting techniques for estimating particle transmission energies, and *a declassified and pioneering article by Enrico Fermi and R.D. Richtmyer in 1948 using mean field type but heuristic like Quantum Monte Carlo methodologies for studying neutron diffusions [266]*. For a detailed account of the applications of mean field particle methods in computational physics we refer the reader to the series of articles of M. Caffarel and his co-authors [92, 93, 94, 526, 527].

Since this period until the mid-1990s, these pioneering studies have been further developed by several mathematicians; to name a few, in alphabetical order, J. Gärtner [279], C. Graham [309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320], B. Jourdain [365], K. Oelschläger [467, 468, 469], Ch. Léonard [411], S. Méléard [446, 447, 448], M. Métivier [451], S. Roelly-Coppoletta [448], T. Shiga and H. Tanaka [537], and A.S. Sznitman [559]. Most of these developments were centered around solving Martingale problems related to the existence of nonlinear Markov chain models, and the description of propagation of chaos type properties of continuous time IPS models, including McKean-Vlasov type diffusion models, reaction diffusion equations, as well as generalized Boltzmann type interacting jump processes. Their traditional applications were essentially restricted to fluid mechanics, chemistry, and condensed matter theory. Some of these application domains are discussed in some detail in the series of articles [119, 120, 134, 332, 419, 136, 440]. The book [377] provides a recent review on this class of nonlinear kinetic models.

Since the mid-1990s, there has been a virtual explosion in the use of mean field IPS methods as a powerful tool in real-word applications of Monte Carlo simulation in information theory, engineering

sciences, numerical physics, and statistical machine learning problems. These sophisticated population type IPS algorithms are also ideally suited to parallel and distributed environment computation [68, 283, 286, 420]. As a result, over the past few years the popularity of these computationally intensive methods has dramatically increased thanks to the availability of cheap and powerful computers. These advanced Monte Carlo integration theories offer nowadays a complementary tool, and a powerful alternative to many standard deterministic function-based projections and deterministic grid-based algorithms, often restricted to low dimensional spaces and linear evolution models.

In contrast to traditional MCMC techniques (including Gibbs sampling techniques [284], which are a particular instance of Metropolis-Hasting models), another central advantage of these mean field IPS models is the fact that their precision parameter is not related to some stationary target measure, nor of some burning time period, but only to *the size of the population*. This precision parameter is more directly related to the computational power of parallel computers on which we are running the IPS algorithms.

Some application domains

In the last two decades, the numerical solving of concrete and complex nonlinear filtering problems, the computation of complex posterior Bayesian distribution, as well as the numerical solving of optimization problems in evolutionary computing, has been revolutionized by this new class of mean field IPS samplers [99, 172, 170, 195, 234, 534, 568]. Nowadays, their range of application is extending from the traditional fluid mechanics modeling towards a variety of nonlinear estimation problems arising in several scientific disciplines; to name a few, with some reference pointers: Advanced signal processing and nonlinear filtering [111, 160, 170, 163, 172, 237, 234, 235, 302, 383, 381], Bayesian analysis and information theory [99, 141, 172, 174, 234, 264, 420], queueing networks [35, 310, 311, 313, 314, 315], control theory [390, 354, 355, 597], combinatorial counting and evolutionary computing [7, 172, 195, 534, 568], image processing [8, 159, 263, 493], data mining [508], molecular and polymer simulation [172, 195, 323], rare events analysis [133, 130, 172, 184, 185, 296], quantum Monte Carlo methods [18, 97, 184, 185, 343, 450, 516], as well as evolutionary algorithms and interacting agent models [85, 166, 300, 344, 534, 568].

Applications on nonlinear filtering problems arising in turbulent fluid mechanics and weather forecasting predictions can also be found in the series of articles by Ch. Baehr and his co-authors [27, 28, 30, 31, 407, 505, 558]. More recent applications of mean field IPS models to spatial point processes, and multiple object filtering theory can be found in the series of articles [108, 109, 110, 145, 233, 476, 477, 478, 530, 586, 587, 588, 589]. These spatial point processes, and related estimation problems occur in a wide variety of scientific disciplines, such as environmental models, including forestry and plant ecology modeling, as well as biology and epidemiology, seismology, materials science, astronomy, queuing theory, and many others. For a detailed discussion on these applications areas we refer the reader to the book of D. Stoyan, W. Kendall, and J. Mecke [553] and the more recent books of P.J. Diggle [229] and A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan [26].

The use of mean field IPS models in mathematical finance is more recent. For instance, using the rare event interpretation of particle methods, R. Carmona, J. P. Fouque, and D. Vestal proposed in [104] an interacting particle algorithm for the computation of the probabilities of simultaneous defaults in large credit portfolios. These developments for credit risk computation were then improved in the recent developments by R. Carmona and S. Crépey [101] and by the author and F. Patras in [196]. Following the particle filtering approach which is already widely used to estimate hidden Markov models, V. Genon-Catalot, T. Jeantheau, and C. Laredo [285] introduced particle methods for the estimation of stochastic volatility models.

More generally, this approach has been applied for filtering nonlinear and non-Gaussian Models by R. Casarin [113], R. Casarin, and C. Trecroci [114]. More recently, M. S. Johannes, N. G. Polson, and J.R. Stroud [362] used a similar approach for filtering latent variables such as the jump times and

sizes in jump diffusion price models. Particle techniques can also be used in financial mathematics to design stochastic optimization algorithms. This version of particle schemes was used by S. Ben Hamida and R. Cont in [49] for providing a new calibration algorithm allowing for the existence of multiple global minima. Finally, in [206, 207], interacting particle methods were used to estimate backward conditional expectations for American option pricing.

For a more thorough discussion on the use of mean field methods in mathematical finance, we refer the reader to the review article [102], in the book [103].

As their name indicates, branching and interacting particle systems are of course directly related to individual based population dynamics models arising in biology and natural evolution theory. A detailed discussion on these topics can be found in the articles [58, 116, 118, 329, 327, 328, 459, 462, 463, 472], and references therein.

In this connection, I also quote the more recent and rapidly developing mean field games theory introduced in the mid-2000s by J.M. Lasry and P.L. Lions in the series of pioneering articles [398, 399, 400, 401]. In this context, fluid particles are replaced by agents or companies that interact mutually in competitive social-economic environments so that to find optimal interacting strategies w.r.t. some reward function.

Applications of game theory with multiple agents systems in biology, economics, and finance are also discussed in the more recent studies by V. Kolokoltsov and his co-authors [378, 379, 380], in the series of articles [10, 50, 86, 87, 393, 412, 466, 569], the ones by P.E. Caines, M. Huang, and R.P. Malhamé [346, 347, 348, 349, 350, 351, 352], as well as in the pioneering article by R. J. Aumann [25]. Finite difference computational methods for solving mean field games Hamilton-Jacobi nonlinear equations can also be found in the recent article by Y. Achdou, F. Camilli, and I. Capuzzo Dolcetta [4].

For a more detailed account on this new branch of games theory, I refer to the seminal Bachelier lecture notes given in 2007-2008 by P.L. Lions at the Collège de France [418], as well as the series of articles [301, 346, 443] and references therein.

The illustrations I have chosen are very often at the crossroad of several seemingly disconnected scientific disciplines, including biology, physics, engineering sciences, probability, and statistics.

In this connection, I emphasize that most of the mean field IPS algorithms I have discussed in these lectures are mathematically identical, but their interpretations strongly depend on the different application domains they are thought. Furthermore, different ways of interpreting a given mean field particle technique often guide researchers' and development engineers' intuition to design and to analyze a variety of consistent mean field stochastic algorithms for solving concrete estimation problems. This variety of interpretations is one of the threads that guide the development of this course, with constant interplays between the theory and the applications.

In fluid mechanics, and computational physics, mean field particle models represent the physical evolution of different kinds of macroscopic quantities interacting with the distribution of microscopic variables. These stochastic models includes physical systems such as gases, macroscopic fluid models, and other molecular chaotic systems. One central modeling idea is often to neglect second order fluctuation terms in complex systems so that to reduce the model to a closed nonlinear evolution equation in distribution spaces (see for instance [134, 547], and references therein). The mean field limit of these particle models represents the evolution of these physical quantities. They are often described by nonlinear integro-differential equations.

In computational biology and population dynamic theory, the mathematical description of mean field genetic type adaptive populations, and related spatial branching processes, is expressed in terms of birth and death and competitive selection type processes, as well as mutation transitions of individuals, also termed particles. The state space of these evolution models depends on the application domain. In genealogical evolution models, the ancestral line of individuals evolves in the space of random trajectories. In genetic population models, individuals are encoded by strings in finite or Euclidian product spaces. Traditionally, these strings represent the chromosomes or the genotypes of the genome. The mutation transitions represent the biological random changes of individuals. The selection process

is associated with fitness functions that evaluate the adaptation level of individuals. In this context, the mean field limit of the particle models is sometimes called the infinite population model. For finite state space models, these evolutions are described by deterministic dynamical systems in some simplex. In more general situations, the limiting evolution belongs to the class of measure valued equations.

In computer sciences, mean field genetic type IPS algorithms (*abbreviated GA*) are also used as random search heuristics that mimic the process of evolution to generate useful solutions to complex optimization problems. In this context, the individuals represent candidate solutions in a problem dependent state space; and the mutation transition is analogous to the biological mutation so as to increase the variability and the diversity of the population of solutions. The selection process is associated with some fitness functions that evaluate the quality of a solution w.r.t. some criteria that depend on the problem at hand. In this context, the limiting mean field model is often given by some Boltzmann-Gibbs measures associated with some fitness potential functions.

In advanced signal processing as well as in statistical machine learning theory, mean field IPS evolution models are also termed Sequential Monte Carlo samplers. As their name indicates, the aim of these methodologies is to sample from a sequence of probability distributions with an increasing complexity on general state spaces, including excursion spaces in rare event level splitting models, transition state spaces in sequential importance sampling models, and path space models in filtering and smoothing problems. In signal processing these evolution algorithms are also called particle filters. In this context, the mutation-selection transitions are often expressed in terms of a prediction step, and the updating of the particle population scheme. In this case, the limiting mean field model coincides with the evolution of conditional distributions of some random process w.r.t. some conditioning event. For linear-Gaussian models, the optimal filter is given by Gaussian conditional distributions, with conditional means and error covariance matrices computed using the celebrated Kalman filter recursions. In this context, these evolution equations can also be interpreted as McKean-Vlasov diffusion type models. The resulting mean field model coincides with the Ensemble Kalman Filters (*abbreviated EKF*) currently used in meteorological forecasting and data assimilation problems.

In physics and molecular chemistry, mean field IPS evolution models are used to simulate quantum systems to estimate ground state energies of a many-body Schödinger evolution equation. In this context, the individuals are termed walkers to avoid confusion with the physical particle based models. These walkers evolve in the set of electronic or macromolecular configurations. These evolution stochastic models belong to the class of Quantum and Diffusion Monte Carlo methods (*abbreviated QMC and DMC*). These Monte Carlo methods are designed to approximate the path space integrals in many-dimensional state spaces. Here again these mean field genetic type techniques are based on a mutation and a selection style transition. During the mutation transition, the walkers evolve randomly and independently in a potential energy landscape on particle configurations. The selection process is associated with a fitness function that reflects the particle absorption in an energy well. In this context, the limiting mean field equation can be interpreted as a normalized Schrödinger type equation. The long time behavior of these nonlinear semigroups is related to top eigenvalues and ground state energies of Schrödinger's operators. A pedagogical introduction to QMC methods can be found in the review articles [93, 94]. For more recent advances including concrete applications in computational physics we refer to the articles [92, 93, 526, 527].

In probability theory, mean field IPS models can be interpreted into two ways. Firstly, the particle scheme can be seen as a step by step projection of the solution of an evolution equation in distribution spaces, into the space of empirical measures. More precisely, the empirical measures associated with a mean field IPS model evolve as a Markov chain in reduced finite dimensional state spaces. In contrast to conventional MCMC models based on the long time behavior of a single stochastic process, the mean field IPS Markov chain model is associated with a population of particles evolving in product state spaces. In this sense, mean field particle methods can be seen as a stochastic linearization of nonlinear equations. The second interpretation of these models relies on an original stochastic perturbation theory of nonlinear evolution equations in distribution spaces. More precisely, the local

sampling transitions of the population of individuals induce some local sampling error, mainly because the transition of each individual depends on the occupation measure of the systems. In this context, the occupation measures of the IPS model evolve as the limiting equation, up to these local sampling errors.

Course outline

The lecture notes cover topics in the general area of Monte Carlo methods and their application domains. The topics include Markov chain Monte Carlo and Sequential Monte Carlo methods, Quantum and Diffusion Monte Carlo techniques, as well as branching and interacting particle methodologies. We cover discrete and continuous time stochastic models, starting from traditional sampling techniques (perfect simulation, Metropolis-Hasting, and Gibbs-Glauber models) to more refined methodologies such as gradient flows diffusions on constraint state space and Riemannian manifolds, ending with the more recent and rapidly developing Branching and mean field type Interacting Particle Systems techniques.

The course offers a pedagogical introduction to the theoretical foundations of these advanced stochastic models, combined with a series of concrete illustrations taken from different application domains. The applications considered in these lectures will range from Bayesian statistical learning (hidden Markov chain, statistical machine learning), risk analysis and rare event sampling (mathematical finance, and industrial risk assessment), operation research (global optimization, combinatorial counting and ranking), advanced signal processing (stochastic nonlinear filtering and control, and data association and multiple objects tracking), computational and statistical physics (molecular dynamics, Schrödinger's ground states, Boltzmann-Gibbs distributions, and free energy computation). Approximately the first half of the course will be concerned with linear type Markov chain Monte Carlo methods, and the second part to nonlinear particle type methodologies.

To illustrate the course, in the first part we present a brief overview of some Monte Carlo methodologies discussed in these lectures:

These stochastic models include standard Monte Carlo methods and conventional Markov chain Monte Carlo techniques based on the law of large numbers and the ergodic theorem. We also present nonlinear mean field type particle methodologies and Feynman-Kac type evolutionary particle techniques, including genealogical tree based models and backward particle Markov chain sampling methods.

These particle methods are illustrated with some selected applications arising in signal processing, information theory and computational physics. Most of the results presented in this introductory chapter are presented without a single proof. We provide precise reference pointers to sections in the lectures dedicated to the modeling and the analysis of these stochastic models.

All the Monte Carlo methodologies presented in this introductory part are sufficiently detailed so that they can be encoded in a computer. We have tried to present in an informal way the mathematical foundations of these techniques and their application domains.

Nevertheless, their rigorous analysis and their performance often depend on sophisticated stochastic models. We encourage the reader to select one of these methodologies and follow the pointers given in the text to answer to the following questions:

- Find the class of stochastic model:

  linear or nonlinear Markov chains, stochastic flow, mean field particle models, ...

- Find the range of applications:

  signal processing, computational physics, operation research, statistical machine learning,....

- Find the precision degree and develop the performance and the convergence of the scheme:

long time behavior, large population sized, double asymptotic, ...

To guide the reader, we use the series of chili logos  , to underline the level of complexity of the mathematical or physical foundations of the stochastic model.

The rest of the lecture notes is decomposed into two parts. The first one is concerned with linear type Monte Carlo methods, the second one is dedicated with nonlinear Markov chain models and their mean field particle interpretations.

Linear Monte Carlo methodologies refer to stochastic methods that can be described by Markov processes. This class of models include discrete and continuous time models, importance sampling methods, Markov and sub-Markov models, Markov chain Monte Carlo methodologies, as well as stochastic processes evolving in constraint manifolds.

Nonlinear Monte Carlo methodologies refer a class of stochastic sampling techniques based on particle interpretations of nonlinear Markov chain models.This class of models include McKean-Vlasov type diffusions, interacting jump processes, and Feynman-Kac path-particle integration models. We illustrate these models with a series of applications arising in computational physics, signal processing, and Bayesian inference.

Pierre Del Moral, Rio de Janeiro 2018

Frequently used notation

Some state spaces

We present some basic notation and background on stochastic analysis and integral operator theory. The lecture notes contain cross-references to this rather well known material; so the reader may wish to skip this section.

We shall use the symbol $a := b$ to define a mathematical object a in terms of b, or vice versa. In these lecture notes, we often use the letters m, n, p, q, k, l to denote integers and r, s, t, u to denote real numbers. We also use the capital letters U, V, WX, Y, Z to denote random variables, and the letters u, v, w, x, y, z the denote their possible outcomes.

We often use the letters E or S to denote some general state space model. To avoid repetition these general state spaces, and all the functions on these spaces are assumed to be measurable; that is, they are equipped with some sigma-field so that the Lebesgue integral is well defined w.r.t. these functions (for instance $\mathbb{R}^d$ equipped with the sigma field generated by the open sets, as well as $\mathbb{N}^d$, $\mathbb{Z}^d$ or any other countable state space equipped with the discrete sigma-field). Given a Polish space E (i.e., a separable and completely metrizable topological space), we denote by $D([a, b], E)$ the set of càdlàg paths from the interval $[a, b]$ into E. The abbreviation càdlàg comes from the French description "continue à droite, limitée à gauche," and the English translation "right continuous with left limits."

We also denote, respectively, by $\mathcal{M}(E)$, $\mathcal{M}_+(E)$, $\mathcal{M}_0(E)$, $\mathcal{P}(E)$, and $\mathcal{B}(E)$, the set of all finite signed measures on some (measurable) space E, the subset of positive measures, the convex subset of measures with null mass, the set of all probability measures, and the Banach space of all bounded and measurable functions f equipped with the uniform norm $\|f\| = \mathrm{Sup}_{x \in E} |f(x)|$.

We also denote by $\mathrm{Osc}(E)$, and by $\mathcal{B}_1(E)$, the set of $\mathcal{E}$-measurable functions f with oscillations $\mathrm{osc}(f) = \mathrm{Sup}_{x,y} |f(x) - f(y)| \leq 1$, and, respectively, with $\|f\| \leq 1$.

We also often use the letters f, g, h or F, G, H to denote functions on some state space, and μ, ν, η or $\mu(dx), \nu(dx), \eta(dx)$ measures on some state space.

Given a pair of functions f_1 and f_2 on some state spaces E_1 and E_2. we denote by $(f_1 \otimes f_2)$ the tensor product function on the product space $(E_1 \times E_2)$ defined for any $(x_1, x_2) \in (E_1 \times E_2)$ by

$$(f_1 \otimes f_2)(x_1, x_2) = f_1(x_1) f_2(x_2)$$

We also use the proportional sign $f \propto g$ between functions to mention that $f = c\ g$ for some constant $c \in \mathbb{R}$.

We shall slightly abuse the notation, denoting by 0 and 1 the zero and the unit elements in the semi-rings $(\mathbb{R}, +, \times)$ and by 0 and 1 the zero and the unit elements in the set of functions on some state space E.

The maximum and minimum operations are denoted respectively by

$$a \vee b := \max\{a, b\} \quad a \wedge b := \min\{a, b\} \quad \text{as well as} \quad a_+ := a \vee 0$$

We also denote by $\lfloor a \rfloor$ and $\{a\} = a - \lfloor a \rfloor$ the integer part, and resp. the fractional part, of some real number a.

We also use the Bachmann-Landau notation

$$f(\epsilon) = g(\epsilon) + \mathrm{O}(\epsilon) \Longleftrightarrow \limsup_{\epsilon \to 0} \frac{1}{\epsilon} |f(\epsilon) - g(\epsilon)| < \infty$$

and

$$f(\epsilon) = g(\epsilon) + \mathrm{o}(\epsilon) \Longleftrightarrow \limsup_{\epsilon \to 0} \frac{1}{\epsilon} |f(\epsilon) - g(\epsilon)| = 0$$

When there are no confusions, sometimes we write $\mathrm{o}(1)$ a function that tends to 0 when the parameter $\epsilon \to 0$. We also denote by $\mathrm{O}_P(\epsilon)$ some possibly random function such that

$$\mathbb{E}\left(|\mathrm{O}_P(\epsilon)|\right) = \mathrm{O}(\epsilon)$$

We also use the traditional conventions

$$\prod_{\emptyset} = 1 \qquad \sum_{\emptyset} = 0 \qquad \inf_{\emptyset} = \infty \qquad \text{and} \qquad \sup_{\emptyset} = -\infty$$

Integration

Given a measure η on some (measurable) state space E and some (measurable) function f from E into $\mathbb{R}$ we set

$$\eta(f) = \int \eta(dx) f(x)$$

as soon as the integral exists, that is if f is integrable w.r.t. η; more formally if $\eta(|f|) < \infty$. For indicator functions $f = 1_A$, sometimes we slightly abuse notation and we set $\eta(A)$ instead of $\eta(1_A)$

$$\eta(1_A) = \int \eta(dx) 1_A(x) = \int_A \eta(dx) = \eta(A)$$

Sometimes, we simplify the presentation, denoting by $d\mu$ the measure $d\mu(x) = \mu(dx)$, where dx stands for an infinitesimal neighborhood of the state $x \in E$.

We also consider the partial order relation between functions f_1, f_2 and measures μ_1, μ_2 given by

$$f_1 \leq f_2 \Longleftrightarrow \forall x \in S \quad f_1(x) \leq f_2(x)$$

and

$$\mu_1 \leq \mu_2 \Longleftrightarrow \forall A \in \mathcal{S} \quad \mu_1(A) \leq \mu_2(A)$$

We denote by $(\mu_1 \otimes \mu_2) \in \mathcal{M}(E_1 \times E_2)$ the tensor product measure defined for any $f \in \mathcal{B}_b(E_1 \times E_2)$ by

$$(\mu_1 \otimes \mu_2)(f) = \int \mu_1(dx_1) \mu_2(dx_2) f(x_1, x_2)$$

The Dirac measure δ_a at some point $a \in S$ is defined by

$$\delta_a(f) = \int f(x) \delta_a(dx) = f(a)$$

When η is the distribution of some r.v. X taking values in S, we have

$$\eta(dx) = \mathbb{P}(X \in dx) \quad \text{and} \quad \eta(f) = \mathbb{E}(f(X))$$

For instance the measure on $\mathbb{R}$ given by

$$\eta(dx) = \frac{1}{2} \left(\frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \, dx \right) + \frac{1}{2} \left(\frac{1}{2} \delta_0(dx) + \frac{1}{2} \delta_1(dx) \right)$$

represents the distribution of the random variable

$$X := \epsilon\ Y + (1-\epsilon)Z$$

where (ϵ, Y, Z) are independent r.v. with distribution

$$\begin{aligned} \mathbb{P}(\epsilon = 1) &= 1 - \mathbb{P}(\epsilon = 0) = 1/2 \\ \mathbb{P}(Z = 1) &= 1 - \mathbb{P}(Z = 0) = 1/2 \quad \text{and} \quad \mathbb{P}(Y \in dy) = \frac{1}{\sqrt{2\pi}}\ e^{-y^2/2}\ dy \end{aligned} \tag{1}$$

We use the notation $dx (= dx_1 \times \ldots \times dx_k)$ to denote the Lebesgue measure on some Euclidian space $\mathbb{R}^k$, of some $k \geq 1$. For finite or countable state spaces, measures are identified to function and sometimes we write $\mu(x), \nu(x), \eta(x)$ instead of $\mu(dx), \nu(dx), \eta(dx)$. Notice that any countable state space E can be embedded in $\mathbb{R}$ so that the measures on E can also be represented by a weighted Dirac measures μ on $\mathbb{R}$ with support in E. In this case, $\mu(\{x\}) = \mu(x)$.

More precisely, for finite spaces of the form $S = \{e_1, \ldots, e_d\} \subset \mathbb{R}$, measures are defined by the weighted Dirac measures

$$\eta = \sum_{1 \leq i \leq d} w_i\ \delta_{e_i} \quad \text{with} \quad w_i = \eta(\{e_i\}) := \eta(e_i)$$

so that

$$\eta(f) = \int \eta(dx) f(x) = \sum_{1 \leq i \leq d} \eta(e_i) f(e_i)$$

Thus, if we identify measures and functions by the line and column vectors

$$\eta = [\eta(e_1), \ldots, \eta(e_d)] \quad \text{and} \quad f = \begin{pmatrix} f(e_1) \\ \vdots \\ f(e_d) \end{pmatrix} \tag{2}$$

we have

$$\eta f = [\eta(e_1), \ldots, \eta(e_d)] \begin{pmatrix} f(e_1) \\ \vdots \\ f(e_d) \end{pmatrix} = \sum_{1 \leq i \leq d} \eta(e_i) f(e_i) = \eta(f) \tag{3}$$

The Dirac measure δ_{e_i} is simply given by the line vector

$$\delta_{e_i} = \left[0, \ldots, 0, \underbrace{1}_{i-th}, 0, \ldots, 0 \right]$$

In this notation, probability measures on S can be interpreted as a point $(\eta(e_i))_{1 \leq i \leq d}$ in the $(d-1)$-dimensional simplex $\Delta_{d-1} \subset [0,1]^d$ defined by

$$\Delta_{d-1} = \left\{ (p_1, \ldots, p_d) \in [0,1]^d \ :\ \sum_{1 \leq i \leq d} p_i = 1 \right\} \tag{4}$$

Integral operators

Finite spaces and matrices

We consider a couple of r.v. (X_1, X_2) on a state space $(E_1 \times E_2)$, with marginal distributions

$$\eta_1(dx_1) = \mathbb{P}(X_1 \in dx_1) \quad \text{and} \quad \eta_2(dx_2) = \mathbb{P}(X_2 \in dx_2)$$

and conditional distribution

$$M(x_1, dx_2) = \mathbb{P}(X_2 \in dx_2 \mid X_1 = x_1)$$

For finite spaces of the form $E_1 = \{a_1, \ldots, a_{d_1}\}$ and $E_2 := \{b_1, \ldots, b_{d_2}\} \subset E = \mathbb{R}$, the above conditional distribution can be represented by a matrix

$$\begin{pmatrix} M(a_1, b_1) & M(a_1, b_2) & \ldots & M(a_1, b_{d_2}) \\ M(a_2, b_1) & M(a_2, b_2) & \ldots & M(a_2, b_{d_2}) \\ \vdots & \vdots & \vdots & \vdots \\ M(a_{d_1}, b_1) & M(a_2, b_2) & \ldots & M(a_{d_2}, b_{d_2}) \end{pmatrix}$$

By construction, we have

$$\underbrace{\mathbb{P}(X_2 \in dx_2)}_{=\eta_2(dx_2)} = \int_{E_1} \underbrace{\mathbb{P}(X_1 \in dx_1)}_{\eta_1(dx_1)} \times \underbrace{\mathbb{P}(X_2 \in dx_2 \mid X_1 = x_1)}_{M(x_1, dx_2)}$$

In other words, we have

$$\eta_2(dx_2) = \int_{S_1} \eta_1(dx_1)\ M(x_1, dx_2) := (\eta_1 M)(dx_2)$$

or in a more synthetic form $\eta_2 = \eta_1 M$. Notice that for the finite state space model discussed above we have the matrix formulation

$$\begin{aligned} \eta_2 &= [\eta_2(b_1), \ldots, \eta_2(b_{d_2})] \\ &= [\eta_1(a_1), \ldots, \eta_1(a_{d_1})] \begin{pmatrix} M(a_1, b_1) & M(a_1, b_2) & \ldots & M(a_1, b_{d_2}) \\ M(a_2, b_1) & M(a_2, b_2) & \ldots & M(a_2, b_{d_2}) \\ \vdots & \vdots & \vdots & \vdots \\ M(a_{d_1}, b_1) & M(a_2, b_2) & \ldots & M(a_{d_2}, b_{d_2}) \end{pmatrix} = \eta_1 M \end{aligned}$$

In this context, a matrix M with positive entries whose rows sum to 1 is also called a stochastic matrix.

Given a function f on E_2, we consider the function $M(f)$ on E_1 defined by

$$M(f)(x_1) = \int_{E_2} M(x_1, dx_2)\ f(x_2) = \mathbb{E}\left(f(X_2) \mid X_1 = x_1\right)$$

Here again, for the finite state space model discussed above these definitions resume to matrix operations

$$\begin{aligned} M(f) &= \begin{pmatrix} M(f)(a_1) \\ \vdots \\ M(f)(a_{d_1}) \end{pmatrix} \\ &= \begin{pmatrix} M(a_1, b_1) & M(a_1, b_2) & \ldots & M(a_1, b_{d_2}) \\ M(a_2, b_1) & M(a_2, b_2) & \ldots & M(a_2, b_{d_2}) \\ \vdots & \vdots & \vdots & \vdots \\ M(a_{d_1}, b_1) & M(a_2, b_2) & \ldots & M(a_{d_2}, b_{d_2}) \end{pmatrix} \begin{pmatrix} f(b_1) \\ \vdots \\ f(b_{d_2}) \end{pmatrix} \end{aligned}$$

By construction, we also have that

$$\eta_1(M(f)) = (\eta_1 M)(f) = \eta_2(f) \Longleftrightarrow \mathbb{E}\left(\mathbb{E}(f(X_2)|X_1)\right) = \mathbb{E}(f(X_2)) \tag{5}$$

Given some matrices M, M_1 and M_2, we denote by $M_1 M_2$ the composition of the matrices M_1 and M_2, and by $M^n = M^{n-1} M = M M^{n-1}$ the n iterates of M. For $n = 0$, we use the convention $M^0 = Id$, the identity matrix on S.

Abstract integral and differential operators

A bounded integral operator Q from a (measurable) space E_1 into an auxiliary (measurable) space E_2 is an operator $f_2 \mapsto f_1 = Q(f_2)$ from $\mathcal{B}(E_2)$ into $\mathcal{B}(E_1)$ such that the functions

$$x_1 \in E_1 \mapsto f_1(x_1) = Q(f_2)(x_1) := \int_{E_2} Q(x_1, dx_2) f_2(x_2)$$

are (measurable) bounded, for any $f_2 \in \mathcal{B}(E_2)$. Depending on its action, the operator Q is alternatively called a bounded integral operator from E_1 into E_2, or from $\mathcal{B}(E_2)$ into $\mathcal{B}(E_1)$.

A positive operator is a bounded integral operator Q, such that $Q(f) \geq 0$, for any $f \geq 0$. A Markov kernel, or a Markov transition, is a positive and bounded integral operator Q with $Q(1) = 1$.

A bounded integral operator Q from a measurable space E_1 into an auxiliary measurable space E_2 also generates a dual operator

$$\mu_1(dx_1) \mapsto \mu_2(dx_2) = (\mu_1 Q)(dx_2) = \int \mu_1(dx_1) Q(x_1, dx_2)$$

from $\mathcal{M}(E_1)$ into $\mathcal{M}(E_2)$ defined by $(\mu_1 Q)(f_2) := \mu_1(Q(f_2))$, for any $f_2 \in \mathcal{B}(E_2)$. Sometimes, with a slight abuse of notation, to the presentation we write $Q(x_1, A_2)$ instead of $Q(1_{A_2})(x_1)$ for some (measurable) subset $A_2 \subset E_2$ and for any $x_1 \in E_1$.

For finite state spaces of the form $E_1 = \{a_1, \ldots, a_{d_1}\}$ and $E_2 := \{b_1, \ldots, b_{d_2}\} \subset E = \mathbb{R}$, using the identification of functions and measures with column and row vectors (2) the operations $f \mapsto Q(f)$ and $\eta \mapsto \eta Q$ reduces to standard matrix vector operations with

$$\begin{pmatrix} Q(a_1, b_1) & Q(a_1, b_2) & \ldots & Q(a_1, b_{d_2}) \\ Q(a_2, b_1) & Q(a_2, b_2) & \ldots & Q(a_2, b_{d_2}) \\ \vdots & \vdots & \vdots & \vdots \\ Q(a_{d_1}, b_1) & Q(a_2, b_2) & \ldots & Q(a_{d_2}, b_{d_2}) \end{pmatrix}$$

Given some measure μ on E_1, sometimes we write $\mu_1 \otimes Q$, the measure on $(E_1 \times E_2)$ defined by

$$[\mu \otimes Q](d(x_1, x_2)) = \mu(dx_1) Q(x_1, dx_2)$$

For Markov transitions $Q = M$, the function $x_1 \mapsto M(f_2)(x_1)$ represents the local averages of f_2 around x_1. More precisely, if $X_2(x_1)$ stands for some random variable with law $Q(x_1, dx_2)$ on E_2 we have

$$M(f_2)(x_1) = \mathbb{E}\left(f(X_2(x_1))\right)$$

In the same way, the mapping $\mu_1 \in \mathcal{P}(E_1) \mapsto \mu_1 M \in \mathcal{P}(E_2)$ represents the distributions of the random states X_2 associated with a given random transition $X_1 \rightsquigarrow X_2$, starting with the distribution $\text{Law}(X_1) = \mu_1$, and sampling a state X_2 with distribution $M(X_1, dx_2)$. More precisely, for any (measurable) subset $A_2 \subset E_2$ we have

$$(\mu_1 M)(A_2) := \mathbb{P}\left(X_2 \in A_2\right) = \int \mathbb{P}(X_1 \in dx_1)\ \mathbb{P}\left(X_2 \in A_2 \,|X_1 = x_1\right) = \int \mu_1(dx_1)\ M(x_1, A_2)$$

In this context, $\mu_1 \otimes M$ represents the distribution of the couple of random variables (X_1, X_2); that is, we have that

$$(\mu_1 \otimes M)(d(x_1, x_2)) = \mathbb{P}(X_1 \in dx_1)\ \mathbb{P}\left(X_2 \in dx_2 \,|X_1 = x_1\right) = \mathbb{P}\left((X_1, X_2) \in d(x_1, x_2)\right)$$

For instance the integral operator given by

$$M(x_1, dx_2) = \frac{1}{2}\ \frac{1}{\sqrt{2\pi}}\ e^{-(x_2 - a(x_1))^2/2}\ dx_2 + \frac{1}{2}\ \left(\frac{1}{2}\ \delta_{b_0(x_1)} + \frac{1}{2}\ \delta_{b_1(x_1)}\right)(dx_2)$$

represents the conditional distribution of the random variable

$$X_2 := \epsilon_1 \ (a(X_1) + Y) + (1 - \epsilon_1) \ b_{\epsilon_2}(X_1)$$

given $X_1 = x_1$, where (ϵ_1, ϵ_2) stands for two copies of the Bernoulli r.v. ϵ and Y is the Gaussian r.v. defined in (1). It is implicitly assumed that $(\epsilon_1, \epsilon_2, Y, X_1)$ are independent.

Given a pair of bounded integral operators (Q_1, Q_2), we let $(Q_1 Q_2)$ are the composition operator defined by $(Q_1 Q_2)(f) = Q_1(Q_2(f))$. For time homogeneous state spaces, we denote by $Q^m = Q^{m-1}Q = QQ^{m-1}$ the m-th composition of a given bounded integral operator Q, with $m \geq 1$.

Given some bounded integral operators $Q_n(x_{n-1}, dx_n)$, from some measurable state space E_{n-1} into a possibly different measurable state space space E_n, we denote by $Q_{p,n}$ the semigroup defined by

$$Q_{p,n} = (Q_{p+1} Q_{p+2} \ldots Q_n) = Q_{p+1} Q_{p+1,n}$$

with the convention $Q_{n,n} = Id$, the identity matrix for $p = n$.

We also consider differential operators L on smooth functions f on $\mathbb{R}^k$, for some $k \geq 1$. For instance, the operator $f \mapsto L(f)$, defined for any $x \in \mathbb{R}^d$ by

$$L(f)(x) = \sum_{1 \leq i \leq d} a^i(x) \left(\partial_{x_i} f\right)(x) + \sum_{1 \leq i,j \leq d} b^{i,j}(x) \left(\partial_{x_i, x_j} f\right)(x)$$

is defined for any twice differentiable function f, and some given functions $x \mapsto (a^i(x), b^{i,j}(x))$, with $1 \leq i, j \leq d$. Given some measure η on $\mathbb{R}^k$, and whenever the function $L(f)$ is integrable w.r.t. $\eta(dx)$, we also set

$$(\eta L)(f) = \eta(L(f)) := \int \eta(dx) \ L(f)(x)$$

For instance, these integrals are well defined as soon as η is a probability measure and L maps functions f with compactly supported derivatives into bounded functions $L(f)$. When η has a density $g(x)$ w.r.t. the Lebesgue measure dx, we notice that

$$(\eta L)(f) = \int g(x) \ L(f)(x) \ dx := \langle g, L(f) \rangle \tag{6}$$

The r.h.s. bracket stands for the duality between $\mathbb{L}_p(\mathbb{R}^k)$ and $\mathbb{L}_q(\mathbb{R}^k)$ with $\frac{1}{p} + \frac{1}{q} = 1$ with $1 < p, q < \infty$; or the inner product of the Hilbert space $\mathbb{L}_2(\mathbb{R}^k)$, as soon as the functions g and $L(f)$ belong to $\mathbb{L}_p(\mathbb{R}^k)$ and $\mathbb{L}_q(\mathbb{R}^k)$.

Last but not least, many models that we consider in this book are defined in terms of collections of Markov transitions K_η from some (measurable) state space E_1 into another E_2, or differential operators L_η (when $E_1 = E_2 = \mathbb{R}^k$) indexed by the set of probability measures η on E_1. To avoid unnecessary repetition of technical abstract conditions, we frame the standing assumption that the Markov transition $\mathcal{K}((x, \eta), dy) = K_\eta(x, dy)$ from $(E_1 \times \mathcal{P}(E_1))$ into E_2 is well defined integral operators, and $L_\eta(f)$ is a measurable function on $\mathbb{R}^k$ for any sufficiently smooth function f on $\mathbb{R}^k$.

The Dirac bras and kets formalism

In theoretical and computational quantum physics, the inner product and more generally dual operators on vector spaces are often represented using a bra-ket formalism introduced in the end of the 1930s by P. Dirac [230] to avoid too sophisticated matrix operations (not so developed and of current use in the beginning of the 20th century).

For finite d-dimensional Euclidian vector spaces $\mathbb{R}^d$ the bras $\prec \alpha \mid$ and the kets $\mid \beta \succ$ are simply given for any $\alpha = (\alpha_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ and $\beta = (\beta_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ row and column

$$\prec \alpha \mid := [\alpha_1, \ldots, \alpha_d] \quad \text{and} \quad \mid \beta \succ := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_d \end{bmatrix} \quad \Rightarrow \prec \alpha \mid\mid \beta \succ = \prec \alpha \mid \beta \succ := \sum_{1 \leq i \leq d} \alpha_i \beta_i$$

In much the same way, the product of bra $\prec \alpha \mid$ with a linear (matrix) operator Q corresponds to the product of the row vector by the matrix

$$\prec \alpha \mid Q := [\alpha_1, \ldots, \alpha_d] \begin{bmatrix} Q(1,1) & Q(1,2) & \ldots & Q(1,d) \\ Q(2,1) & Q(2,2) & \ldots & Q(2,d) \\ \vdots & \vdots & \ldots & \vdots \\ Q(d,1) & Q(d,2) & \ldots & Q(d,d) \end{bmatrix}$$

Likewise, the product of a linear (matrix) operator Q with a ket $\mid \beta \succ$ corresponds to the product of the matrix by the column vector

$$Q \mid \beta \succ := \begin{bmatrix} Q(1,1) & Q(1,2) & \ldots & Q(1,d) \\ Q(2,1) & Q(2,2) & \ldots & Q(2,d) \\ \vdots & \vdots & \ldots & \vdots \\ Q(d,1) & Q(d,2) & \ldots & Q(d,d) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}$$

Combining these operations, we find that

$$\prec \alpha \mid Q \mid \beta \succ = [\alpha_1, \ldots, \alpha_d] \begin{bmatrix} Q(1,1) & Q(1,2) & \ldots & Q(1,d) \\ Q(2,1) & Q(2,2) & \ldots & Q(2,d) \\ \vdots & \vdots & \ldots & \vdots \\ Q(d,1) & Q(d,2) & \ldots & Q(d,d) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}$$

Using the vector representation (2) of functions f and measures η on finite states spaces $S = \{e_1, \ldots, e_d\}$, the duality formula between functions and measures (3) takes the following form

$$\prec \eta \mid f \succ = \eta f = \sum_{1 \leq i \leq d} \eta(e_i) f(e_i) := \eta(f)$$

Likewise, for any Markov transition M from $E_1 = \{a_1, \ldots, a_{d_1}\}$ into $E_2 := \{b_1, \ldots, b_{d_2}\}$, and function f on E_2 and any measure η_1 on E_1, the formula (5) takes the form

$$\prec \eta_1 \mid M \mid f \succ = \prec \eta_1 \mid Mf \succ = \eta_1(Mf) = (\eta_1 M)\, f = \prec \eta_1 M \mid f \succ$$

The bra-ket formalism is extended to differential operations discussed in (6) by setting

$$\prec g \mid L \mid f \succ = \int g(x)\ L(f)(x)\ dx := \langle g, L(f) \rangle$$

Boltzmann-Gibbs transformations

Given a positive and bounded potential function G on E, we also denote by Ψ_G the Boltzmann-Gibbs mapping from $\mathcal{P}(E)$ into itself defined for any $\mu \in \mathcal{P}(E)$ by

$$\Psi_G(\mu)(dx) = \frac{1}{\mu(G)}\ G(x)\ \mu(dx) \tag{7}$$

We illustrate this abstract definition with a couple of examples. We let $\mu = \mathrm{Law}(X)$ be the distribution of some r.v. X and $G = 1_A$ the indicator function of some subset $A \subset S$. In this situation, we have

$$\Psi_G(\mu)(f) = \frac{\mu(Gf)}{\mu(G)} = \frac{\mathbb{E}(f(X)1_A(X))}{\mathbb{E}(1_A(X))} = \mathbb{E}\left(f(X) \mid X \in A\right)$$

In other words, we have

$$\Psi_{1_A}(\mu) = \mathrm{Law}(X \mid X \in A)$$

Let (X, Y) be a couple of r.v. with probability density $p(x, y)$ on $\mathbb{R}^{d+d'}$. With a slight abuse of notation, we recall that the conditional density $p(x|y)$ of X given Y is given by the Bayes' formula

$$p(x|y) = \frac{1}{p(y)}\ p(y|x)\ p(x) \quad p(y) = \int p(y|x)\ p(x)\ dx$$

In other words, we have that

$$\mu(dx) = p(x)dx \quad G_y(x) = p(y|x) \Rightarrow \Psi_{G_y}(\mu) = p(x|y)\ dx$$

There is no loss of generality to assume that G is a $]0, 1]$-valued function. For $[0, 1]$-valued potential functions, the transformation is only defined on measures μ s.t. $\mu(G) > 0$. Boltzmann-Gibbs mappings can be interpreted as a nonlinear Markov transport model

$$\Psi_G(\mu) = \mu S_{\mu,G} \tag{8}$$

for some Markov transitions $S_{\mu,G}$ from E into itself. Next we provide three different types of models of current use in the further development of these lecture notes.

Firstly, for $[0, 1]$-valued potential functions G s.t. $\mu(G) > 0$, we can choose

$$S_{\mu,G}(x, dy) := G(x)\ \delta_x(dy) + (1 - G(x))\ \Psi_G(\mu)(dy) \tag{9}$$

For $[1, \infty[$-valued potential functions G s.t. $\mu(G) > 1$, we can also choose the Markov transitions

$$S_{\mu,G}(x, dy) := \frac{1}{\mu\,(G)}\ \delta_x(dy) + \left(1 - \frac{1}{\mu\,(G)}\right)\ \Psi_{(G-1)}(\mu)(dy)$$

For any bounded positive functions G, the Equation (8) is also met for

$$S_{\mu,G}(x, dy) := \epsilon_\mu G(x)\ \delta_x(dy) + (1 - \epsilon_\mu G(x))\ \Psi_G(\mu)(dy) \tag{10}$$

for any $\epsilon_\mu \geq 0$ s.t. $\epsilon_\eta G(x) \leq 1$ for μ-almost every state x. For instance, we can choose $\epsilon_\mu = 0$, $\epsilon_\mu = 1/\|G\|$, or preferably $\epsilon_\mu = 1/\mu - ess\sup G$, where $\mu - ess\sup G$ stands for the μ-essential supremum of G.

Norms and ergodic constants

We say that a measure μ is absolutely continuous w.r.t. another measure ν on E, and we write $\mu \ll \nu$, when we have $\nu(A) = 0 \Rightarrow \mu(A) = 0$, for any $A \in \mathcal{E}$. When $\mu \ll \nu$, we denote by $d\mu/d\nu$ the Radon-Nykodim derivative function.

The total variation distance $\|\mu_1 - \mu_2\|_{\mathrm{tv}}$ between two probability measures $\mu_1, \mu_2 \in \mathcal{P}(E)$ is defined by

$$\begin{aligned}
\|\mu_1 - \mu_2\|_{\mathrm{tv}} &= \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)| \\
&= 2^{-1} \sup \{|\mu_1(f) - \mu_2(f)|\,; \quad \|f\| \leq 1\} = \sup \{|\mu_1(f) - \mu_2(f)|\,; \quad \mathrm{osc}(f) \leq 1\}
\end{aligned}$$

When the bounded integral operator $Q(x_1, dx_2)$ from E_1 into E_2 has a constant mass, that is, when $Q(1)\,(x_1) = Q(1)\,(y_1)$ for any $(x_1, y_1) \in E_1^2$, the operator $\mu \mapsto \mu Q$ maps $\mathcal{M}_0(E_1)$ into $\mathcal{M}_0(E_2)$. In this situation, we let $\beta(Q)$ be the Dobrushin coefficient of a bounded integral operator Q defined by the formula

$$\beta(Q) := \sup\ \{\mathrm{osc}(Q(f))\ ;\ \ f \in \mathrm{Osc}(E_2)\} \tag{11}$$

The Boltzmann relative entropy between a couple of measures $\mu \ll \nu$ on E is the nonnegative distance-like criteria defined by

$$\mathrm{Ent}\,(\mu|\nu) := \mu\left(\log \frac{d\mu}{d\nu}\right) \quad \text{if} \quad \mu \ll \nu$$

If $\mu \not\ll \nu$ then we set $\mathrm{Ent}\,(\mu|\nu) = \infty$.

The V-norm on the set of signed measure $\mathcal{M}(S)$ (on some state space S equipped with some σ-field $\mathcal{S}$) associated with some non negative function V is defined for any $\mu \in \mathcal{M}(S)$ by

$$\|\mu\|_V := \|\mu\|_{tv} + |\mu|\,(V) \tag{12}$$

In the above display, $|\mu| = \mu^+ + \mu^-$ stands for the total variation of the measure μ, defined in terms of the Hahn-Jordan decomposition $\mu = \mu^+ - \mu^-$ of μ.

We define the V-norm and the V-oscillation of a given measurable function f on S by

$$\|f\|_V := \left\|\frac{f}{V+1/2}\right\| = \sup_{x\in S}\left(\frac{|f(x)|}{V(x)+1/2}\right)$$

and

$$\mathrm{osc}_V(f) := \sup_{x,y\in S}\left(\frac{|f(x)-f(y)|}{[V(x)+V(y)+1]}\right) \tag{13}$$

We recall that

$$\|\mu\|_V \;=\; \sup\{|\mu(f)| \;:\; f \text{ s.t. } \mathrm{osc}_V(f) \le 1\} = \sup\{|\mu(f)| \;:\; f \text{ s.t. } \|f\|_V \le 1\} \tag{14}$$

When $V = 0$ we have

$$\mathrm{osc}_0(f) = \mathrm{osc}(f) \quad \text{and} \quad \|f\|_0 = 2\,\|f\|$$

Sometimes these V-norms type quantities are expressed in terms of the functions $W = 1/2 + V(x)$ with

$$\|\boldsymbol{f}\|_{\mathbf{W}} := \sup_{x\in S}\frac{|f(x)|}{W(x)} \quad \text{and} \quad \mathbf{osc_W(\boldsymbol{f})} := \sup_{x,y\in S}\frac{|f(x)-f(y)|}{W(x)+W(y)}$$

Contents

Introduction

Part I

A brief overview

Chapter 1

Finite state space models

1.1 Markov chains and Matrices

Suppose we are given a matrix $(M(i,j))_{1\leq i,j\leq d}$ with positive entries s.t.

$$\forall 1\leq i\leq d \qquad \sum_{1\leq j\leq d} M(i,j)=1$$

These matrices are called stochastic matrices, or Markov chain (elementary) transitions.

For any index $1\leq i\leq d$ and any integer $n\geq 0$, we let $F_n(i)$ be collection of independent and identically distributed random variables with common distribution

$$\forall 1\leq j\leq d \qquad \mathbb{P}\left(F_n(i)=j\right)=M(i,j) \tag{1.1}$$

We start with some initial state $X_0=i$ and we define a sequence of random variables X_n taking values on $E=\{1,\dots,d\}$ by setting

$$X_n=F_n\left(X_{n-1}\right) \tag{1.2}$$

This sequence is called a Markov chain taking values in $E=\{1,\dots,d\}$ with transition probabilities M. In this situation, we have

$$\mathbb{P}\left(X_n=j \mid X_0=i\right)=M^n(i,j)$$

where M^n stands for the n-th iterate of the matrix M. We set

$$\eta_n(j):=\mathbb{P}\left(X_n=j\right) \quad \text{and} \quad \eta_n=[\eta_n(1),\dots,\eta_n(d)]$$

Since any $f\in\mathbb{R}^E=\mathbb{R}^d$ can be represented by a column vector, by construction for any $k\leq n$ we find that

$$\eta_n=\eta_k M^{n-k} \quad \text{and} \quad (M^{n-k}f)(i)=\mathbb{E}\left(f(X_n)\mid X_k=i\right)$$

1.2 Monte Carlo and Markov chain methods

Sampling N independent copies $(X_n^k)_{1\leq k\leq N}$ of the chain X_n defined in (1.2) and starting at $X_0=i$, we have

$$\forall 1\leq j\leq d \qquad \eta_n^N(j):=\frac{1}{N}\sum_{1\leq k\leq N} 1_{X_n^k=j} \;\simeq_{N\uparrow\infty}\; M^n(i,j)=\eta_n(j)$$

More generally, for any function f we have

$$\eta_n^N(f):=\frac{1}{N}\sum_{1\leq k\leq N} f(X_n^k) \;\simeq_{N\uparrow\infty}\; \eta_n(f)$$

In addition, we have the unbiased and variance formula

$$\mathbb{E}\left(\eta_n^N(f)\right) = \eta_n(f) \quad \text{and} \quad \mathbb{E}\left(\left[\eta_n^N(f) - \eta_n(f)\right]^2\right) = \frac{1}{N}\,\eta_n([f-\eta_n(f)]^2)$$

We further assume that $M^m(i,j) > 0$ for some integer $m \geq 1$. In this situation, there exists a single line vector

$$\pi = [\pi(1), \ldots, \pi(d)] \quad \text{s.t.} \quad \sum_{1\leq i\leq d} \pi(i) = 1 \quad \text{and} \quad \pi M = \pi$$

We can also prove that for any function f s.t. $\mathrm{osc}(f) \leq 1$ we have

$$\begin{aligned} \left|\mathbb{E}\left(\eta_n^N(f)\right) - \pi(f)\right| &= |\eta_n(f) - \pi(f)| \leq c^2\, e^{-\alpha n} \\ \mathbb{E}\left(\left[\eta_n^N(f) - \pi(f)\right]^2\right) &= \underbrace{\frac{1}{N}\,\eta_n([f-\eta_n(f)]^2)}_{\text{fluctuation term}} + \underbrace{(\eta_n(f) - \pi(f))^2}_{\text{bias term}} \leq \frac{1}{N} + c^2\, e^{-2\alpha n} \end{aligned}$$

for some finite positive constants c and α. The r.h.s. estimate are direct consequences of the law of large numbers variance formulae combined with exponential stability properties of Markov chain discussed in section 4.4. Assuming the initial state of the chain has distribution π these estimates reduce to

$$\mathbb{E}\left(\eta_n^N(f)\right) = \pi(f) \quad \text{and} \quad \mathbb{E}\left(\left[\eta_n^N(f) - \pi(f)\right]^2\right) = \frac{1}{N}\,\pi([f-\pi(f)]^2)$$

Another way of estimating this vector amounts to count the proportion of times the chain X_n visits each of the states

$$\forall 1 \leq j \leq d \qquad \pi^n(j) := \frac{1}{n}\sum_{0\leq k<n} 1_{X_k=j} \;\simeq_{n\uparrow\infty}\; \pi(j) \tag{1.3}$$

In this situation, assuming that the initial state of the chain has distribution π and using the ergodic analysis developed in section 4.6, we can also prove that for any function f s.t. $\mathrm{osc}(f) \leq 1$ and $\pi(f) = 0$ we have

$$\begin{aligned} \mathbb{E}\left(\pi^n(f)\right) &= \pi(f) \\ \mathbb{E}\left(\left[\pi^n(f) - \pi(f)\right]^2\right) &= \frac{1}{n}\left[\pi([f-\pi(f)]^2) + 2\sum_{p\geq 1}\pi([f-\pi(f)]M^p[f-\pi(f)])\right] \end{aligned}$$

1.3 A couple of Markov chain Monte Carlo models

1.3.1 The Metropolis-Hasting algorithm

Suppose we are given some line vector π such that $\wedge_{i\in E}\pi(i) > 0$ and $\sum_{1\leq i\leq d}\pi(i) = 1$. Choose any Markov transition $M(i,j)$ s.t. $M(i,j) > 0 \Leftrightarrow M(j,i) > 0$ and set

$$a(i,j) = 1 \wedge \frac{\pi(j)M(j,i)}{\pi(i)M(i,j)} \;\in\; [0,1]$$

We let $(U_n)_{n\geq 0}$ be a sequence of independent and identically distributed (*abbreviated i.i.d.*) random variables (*abbreviated r.v.*) on $[0,1]$. We slightly change the Markov chain model (1.2) by introducing an acceptance-rejection step.

$$X_{n-1} \rightsquigarrow Y_n = F_n(X_{n-1}) \rightsquigarrow X_n := \left\{ \begin{array}{rcl} Y_n & \text{if} & U_n \leq a\,(X_{n-1}, Y_n) \\ X_{n-1} & \text{if} & U_n > a\,(X_{n-1}, Y_n) \end{array} \right. \tag{1.4}$$

In this situation, we also have the convergence property (1.3). In addition, if we start with a r.v. with distribution π, all the random states of the chain have the same law π and the law of any random sequence $(X_0, X_1, \ldots, X_n)$ is the same as the law of the reverse trajectory $(X_n, X_{n-1}, \ldots, X_0)$

This Markov chain is called the Metropolis-Hasting sampler and it is discussed in section 7.2.

1.3.2 The Gibbs-Glauber sampler

We end this section with a Markov chain model for sampling a product measure. Suppose we are given a couple of random variables $X = (Y, Z)$ on the product space E^2. We recall that

$$\mathbb{P}((Y,Z)=(i,j)) = \underbrace{\mathbb{P}(Y=i \mid Z=j)}_{=M_1(i,j)}\ \mathbb{P}(Z=j) = \underbrace{\mathbb{P}(Z=j \mid Y=i)}_{=M_2(i,j)}\ \mathbb{P}(Y=i)$$

For any index $1 \leq i \leq d$ and any integer $n \geq 0$, we let $F_n^1(i)$, resp. $F_n^2(i)$ be collection of independent and identically distributed random variables with common distribution

$$\mathbb{P}\left(F_n^1(i)=j\right) = M_1(i,j) \quad \text{and} \quad \mathbb{P}\left(F_n^2(i)=j\right) = M_2(i,j)$$

We consider the Markov chain sequence

$$X_n = \begin{pmatrix} Y_n \\ Z_n \end{pmatrix} \rightsquigarrow X_{n+\frac{1}{2}} = \begin{pmatrix} Y_{n+\frac{1}{2}} \\ Z_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} F_{n+1}^1(Z_n) \\ Z_n \end{pmatrix} \rightsquigarrow X_{n+1} = \begin{pmatrix} Y_{n+\frac{1}{2}} \\ F_{n+1}^2\left(Y_{n+\frac{1}{2}}\right) \end{pmatrix}$$

starting as some state $X_0 = \begin{pmatrix} i \\ j \end{pmatrix}$. In this situation,

$$\forall 1 \leq i,j \leq d \qquad \frac{1}{n}\sum_{0\leq k<n} 1_{(Y_k,Z_k)=(i,j)} \simeq_{n\uparrow\infty} \mathbb{P}((Y,Z)=(i,j))$$

This Markov chain model is called the Gibbs sampler or the Gibbs-Glauber dynamics. This model is studied in some details in section 7.3.

1.4 Interacting Markov Chain Monte Carlo samplers

We further assume that $M(i,j) = M(j,i)$. Suppose we are given some function $V : E \to \mathbb{R}$. We let β_n be some increasing sequence of real numbers, and we set

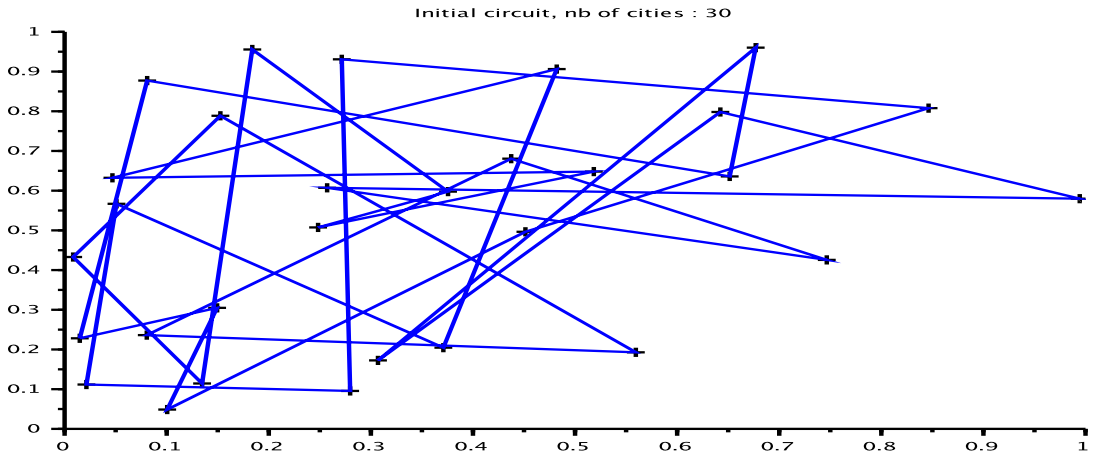
$$a_n(i,j) = 1 \wedge \left[e^{-\beta_n(V(j)-V(i))}\right] \in [0,1] \tag{1.5}$$

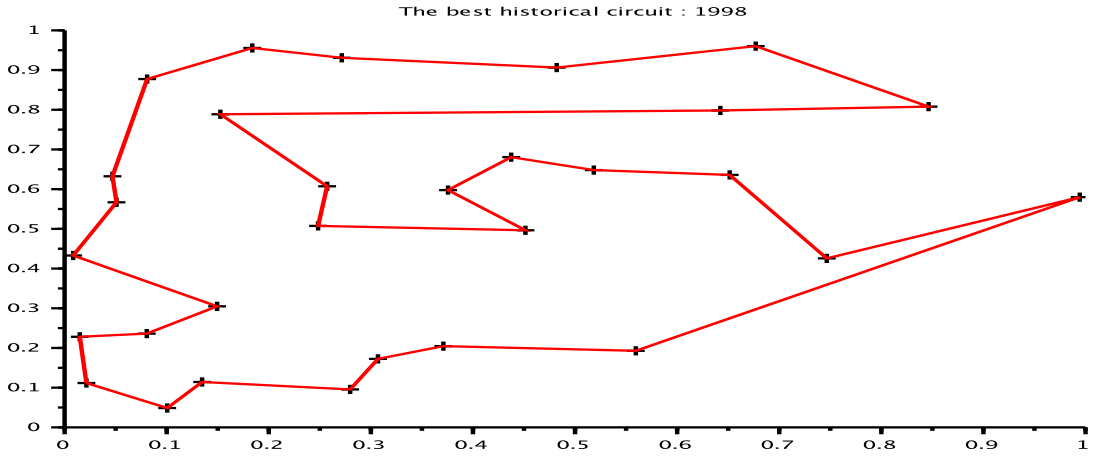
We let X_n be the Markov chain defined as in (1.4) by replacing a by a_n. In this situation, there exists some judicious ways of increasing β_n w.r.t. time so that

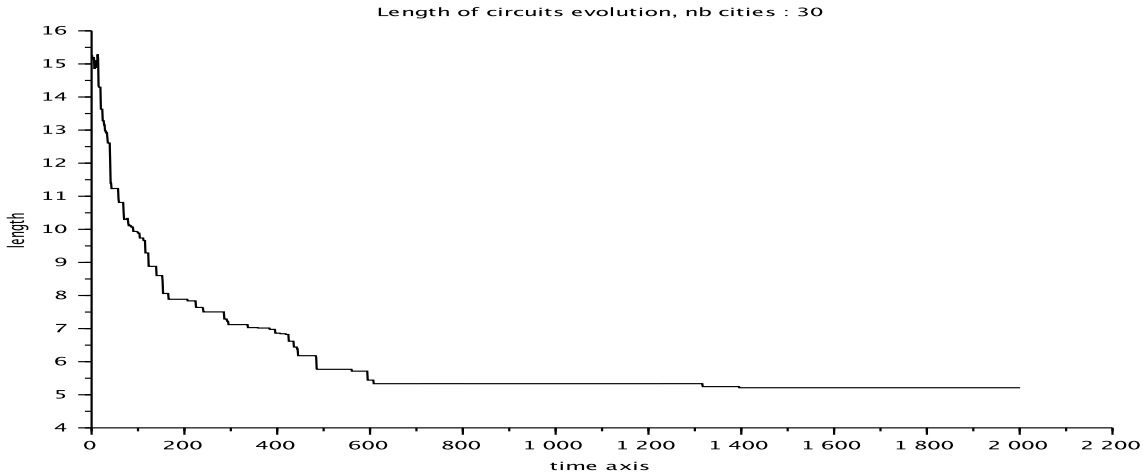
$$V(X_n) \simeq_{n\uparrow\infty} \min_{i\in E} V(i) := V^\star$$

This Markov chain model is called the simulated annealing. Further details on this global optimization algorithm can be found in section 7.5.1.

The following pictures illustrates the initial circuit, the best historical circuit, and the evolution of the length of the circuits in the traveling salesman problem (TSP) with 30 cities.







One natural way to bypass the slow convergence to equilibrium of the simulated annealing between inverse temperature β_n variations is to consider a sequence of interacting simulated annealing models $(\xi_n^i)_{1\leq i\leq N}$ on the product space E^N. These models evolve as a genetic type mutation-selection transition.

$$\xi_n := (\xi_n^i)_{1\leq i\leq N} \xrightarrow{\text{selection}} \widehat{\xi}_n := (\widehat{\xi}_n^i)_{1\leq i\leq N} \xrightarrow{\text{mutation}} \xi_{n+1} \tag{1.6}$$

- During the selection stage, we select the particles ξ_n^i which are better adapted to the change of temperature $\beta_{n+1} = (\beta_{n+1} - \beta_n) + \beta_n$. More precisely, se sample a sequence of i.i.d. uniform r.v. U_i on $[0,1]$, and for each i we set

$$\widehat{\xi}_n = \left\{ \begin{array}{lll} \xi_n^i & \text{if} & U_i \leq e^{-(\beta_{n+1}-\beta_n)V(\xi_n^i)} \\ \widetilde{\xi}_n^i & \text{if} & U_i > e^{-(\beta_{n+1}-\beta_n)V(\xi_n^i)} \end{array} \right.$$

  where $\widetilde{\xi}_n^i$ stands for a sequence of i.i.d. r.v. with distribution

$$\sum_{1\leq k\leq N} \frac{e^{-(\beta_{n+1}-\beta_n)V(\xi_n^k)}}{\sum_{1\leq l\leq N} e^{-(\beta_{n+1}-\beta_n)V(\xi_n^l)}}\ \delta_{\xi_n^i}$$

- During the mutation stage, each particles $\widehat{\xi}_n^i \rightsquigarrow \xi_{n+1}^i$ moves independently according to the simulated annealing transition (1.4) with the acceptance rate (1.5) associated with the inverse temperature β_{n+1}

In this situation, there exists some $\delta > 0$ s.t. for any n we have

$$\frac{1}{N}\sum_{1\leq j\leq N} 1_{V(\xi_n^j)=V^\star} \simeq_{N\uparrow\infty} \frac{1}{\mathrm{Card}(\{i\ :\ V(i)=V^\star\})} + \mathrm{O}\left(e^{-\delta\beta_n}\right)$$

The interacting simulated annealing model presented above belongs to the class of Markov Chain Monte Carlo methods with recycling discussed in section 9.1.5

1.5 Killing and sub-Markov chain models

1.5.1 Feynman-Kac formulae

We let G_n be some $[0,1]$-valued function on a finite space $E = \{1,\ldots,d\}$. We let $(U_n)_{n\geq 0}$ be a sequence of i.i.d. r.v. on $[0,1]$, and $F_n(i)$ the r.v. defined in (1.1). We let c be some cemetery state. We consider the Markov chain on $E \cup \{c\}$

$$X_n^c \rightsquigarrow \widehat{X}_n^c = \left\{ \begin{array}{lll} X_n^c & \text{if} & U_n \leq G_n(X_n^c) \\ c & \text{if} & U_n > G_n(X_n^c) \end{array} \right. \rightsquigarrow X_{n+1}^c = \left\{ \begin{array}{lll} F_{n+1}\left(\widehat{X}_n^c\right) & \text{if} & \widehat{X}_n^c \neq c \\ c & \text{if} & \widehat{X}_n^c = c \end{array} \right.$$

We denote by T the first time n s.t. $\widehat{X}_n^c = c$. In this notation, we have

$$\mathbb{P}\left(X_0^c = i_0,\ldots,X_n^c = i_n\ ;\ T\geq n\right) = \left\{\prod_{0\leq k<n} G_k(i_k)\right\}\ \mathbb{P}\left(X_0 = i_0,\ldots,X_n = i_n\right)$$

and

$$\mathbb{P}\left(T > n\right) = \sum_{(i_0,\ldots,i_n)\in E^n} \left\{\prod_{0\leq k\leq n} G_k(i_k)\right\}\ \mathbb{P}\left(X_0 = i_0,\ldots,X_n = i_n\right)$$

The weighted product probability measures on the r.h.s. are called Feynman-Kac measures. These models are studied in some details in section 4.2.

For time homogenous models $G_n = G$ we notice that the matrix $Q(i,j) = G(i)M(i,j)$ is sub-Markovian, in the sense that

$$\sum_{j\in E} Q(i,j) = G(i) \leq 1$$

Inversely, any sub-Markovian matrix Q satisfying the above condition can be rewritten as

$$Q(i,j) = G(i)M(i,j) \quad \text{with} \quad G(i) = \sum_{j\in E} Q(i,j) \quad \text{and} \quad M(i,j) = Q(i,j)/\sum_{k\in E} Q(i,k)$$

Whenever $Q(i,j) = Q(j,i)$, we have

$$\frac{1}{n}\log \mathbb{P}\left(T > n\right) \simeq_{n\uparrow\infty} \log\lambda \quad \text{and} \quad \mathbb{P}\left(X_n^c = i \mid T \geq n\right) \simeq_{n\uparrow\infty} h(i)/\sum_{j\in E} h(j) \tag{1.7}$$

where λ is the largest eigenvalue of the matrix Q and h the corresponding eigenvector. These sub-Markov models are discussed in section 4.3.

For indicator type functions $G_n = 1_A$ we have

$$\mathbb{P}\left(X_0^c = i_0, \ldots, X_n^c = i_n \mid T > n\right) = \mathbb{P}\left(X_0 = i_0, \ldots, X_n = i_n \mid X_0 \in A, \ldots,\ X_n \in A\right)$$

and

$$\mathbb{P}\left(T > n\right) = \mathbb{P}\left(X_0 \in A, \ldots,\ X_n \in A\right)$$

These killing/absorption models can be extended without further work to more general state spaces. For instance, we can replace X_n by the simple random walk on $E = \mathbb{Z}$ starting at the origin, and set $A = [-7, 7]$. For $n = 31.56\ 10^6$ seconds, the above quantities can be interpreted as the conditional probability of a random path of a simple random walk with elementary transitions per second staying in the interval $[-7, 7]$ in one year.

1.5.2 An acceptance-rejection technique with recycling

One natural way of sampling conditional distributions discussed in the end of section 1.5 is to use a Markov chain $\xi_n := (\xi_n^i)_{1\leq i\leq N}$ evolving on the product spaces E^N, starting at the origin with a genetic type mutation-selection transition (1.6). In this context the elementary transitions are defined as follows:

- During the selection stage, we quote p_n^N the proportion of particles ξ_n^i in $[-7, 7]$. When $p_n^N \neq 0$ each particle ξ_n^i outside the set $[-7, 7]$ is killed and replaced by selecting randomly a successful particle $[-7, 7]$. At the end of this procedure we have $\widehat{\xi}_n := (\widehat{\xi_n^i})_{1\leq i\leq N}$ particles in $[-7, 7]$ (When all particles are outside the desired set, the algorithm need to be restarted).
- During the mutation stage, each particles $\widehat{\xi_n^i}$ moves independently according to the simple random walk transition.

The Markov chain ξ_n belongs to the class of mean field type Feyman-Kac particle models discussed in section 10. In computational physics, these particle models are also termed Resampled Monte Carlo schemes, of reconfiguration walker evolution models. These Monte Carlo methodologies belongs to the class of Quantum Monte Carlo methods.

In this situation, we have the unbiasedness property

$$\mathbb{E}\left(\prod_{0\leq k\leq n} p_k^N\right) = \mathbb{P}\left(X_0 \in A, \ldots,\ X_n \in A\right)$$

as well as

$$\frac{1}{n}\sum_{0\leq k\leq n} \log p_k^N \simeq_{n\uparrow\infty} \lambda \quad \text{and} \quad \frac{1}{n}\sum_{0\leq k\leq n}\frac{1}{N}\sum_{1\leq j\leq N} 1_{\xi_k^j = i} \simeq_{n\uparrow\infty} h(i)/\sum_{j\in E} h(j)$$

In addition, for any $n \geq 0$, and $i \in [-7,7]$, we have

$$\frac{1}{N} \sum_{1 \leq j \leq N} 1_{\xi_n^j = i} \simeq_{N \uparrow \infty} \mathbb{P}\left(X_n = i \mid X_k \in [-7,7] \; k < n\right)$$

Running backwards in time we may trace the whole ancestral line of the j-th particle in $[-7,7]$

$$\xi_{0,n}^j \longleftarrow \xi_{1,n}^j \longleftarrow \ldots \longleftarrow \xi_{n-1,n}^j \longleftarrow \xi_{n,n}^j = \xi_n^j$$

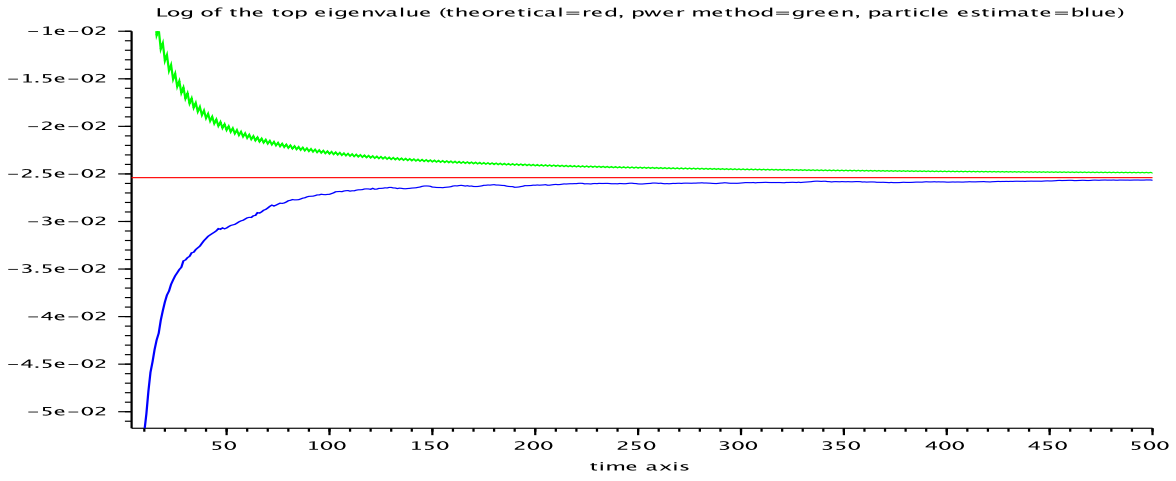
In this notation, for any $n \geq 0$ and any $(i_0, \ldots, i_n) \in [-7,7]^{n+1}$ we have

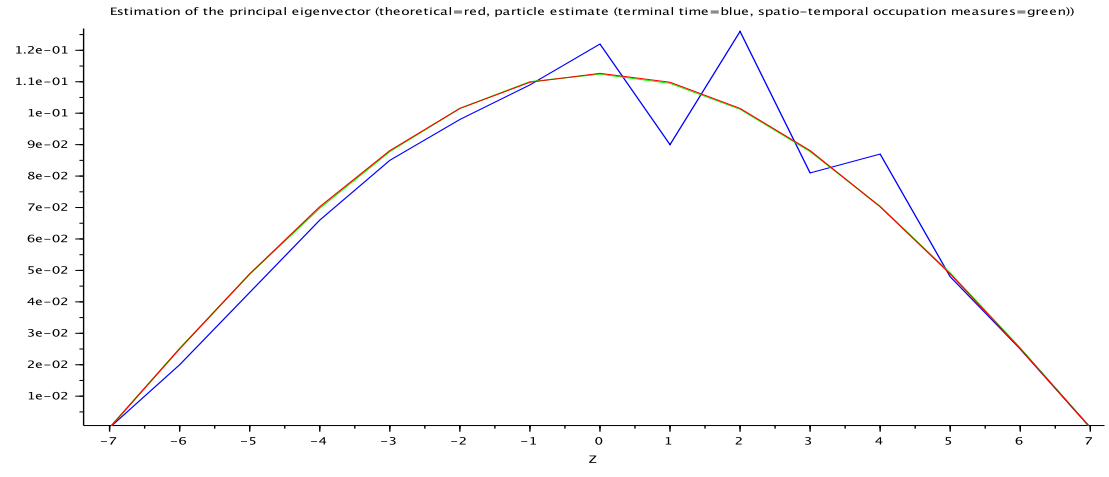
$$\frac{1}{N} \sum_{1 \leq j \leq N} 1_{\xi_{0,n}^j = i_0, \ldots, \xi_{n,n}^j = i_n} \simeq_{N \uparrow \infty} \mathbb{P}\left(X_0 = i_0, \ldots, X_n = i_n \mid X_k \in [-7,7] \; k < n\right)$$

In addition, we have the unbiasedness property

$$\mathbb{E}\left(\left\{\prod_{0 \leq k \leq n} p_k^N\right\} \frac{1}{N} \sum_{1 \leq j \leq N} 1_{\xi_{0,n}^j = i_0, \ldots, \xi_{n,n}^j = i_n}\right) = \mathbb{P}\left(X_0 = i_0, \ldots, X_n = i_n \; ; \; X_k \in [-7,7] \; k < n\right)$$

The next pictures illustrate the estimation of the log of the top eigenvalue and the corresponding eigenvector (1.7) associated with the simple random walk in a tube $[-7,7]$ using the power method (iterates of matrices), and the particle scheme described above.





We fix the time horizon n, and we set

$$\mathbb{X}_0 = \left(\xi^1_{0,n}, \xi^1_{1,n}, \ldots, \xi^1_{n-1,n}, \xi^1_{n,n}\right) \tag{1.8}$$

We sample a new genetic type model $\zeta_n = \left(\zeta^i_n\right)_{1\leq i\leq N}$ defined as above with a frozen trajectory (1.8), in the sense that killed particles at some time k can also be replaced by selecting randomly $\xi^1_{k,n}$ (or another successful particle $[-7,7]$). Then, we choose randomly a genealogical line

$$\mathbb{X}_1 = \left(\xi^I_{0,n}, \xi^I_{1,n}, \ldots, \xi^I_{0,n}, \xi^I_{n,n}\right) \tag{1.9}$$

where I stands for an uniform random variable on $\{1,\ldots,N\}$. Iterating the procedure, we define a Markov chain $(\mathbb{X}_k)_{k\geq 0}$ taking values in E^{n+1}.

In addition, for any $n \geq 0$ and any $(i_0,\ldots,i_n) \in [-7,7]^{n+1}$ we have

$$\frac{1}{m}\sum_{0\leq k<m} 1_{\mathbb{X}_k=(i_0,\ldots,i_n)} \simeq_{m\uparrow\infty} \mathbb{P}\left((X_0,\ldots,X_n)=(i_0,\ldots,i_n) \mid X_k\in[-7,7]\ k<n\right) \tag{1.10}$$

The Markov chain $\mathbb{X}_n$ belongs to the class of Particle Markov Chain Monte Carlo models discussed in section 10.5.2.

Chapter 2

Feynman-Kac particle methods

2.1 Feynman-Kac formulae

The Feynman-Kac models on finite spaces discussed in section 1.5.1 can be extended with a little extra work to more general state space E_n that may even depend on the time parameter n, so that to include path-space Markov chains of given length n. For instance the Feynman-Kac model associated with a collection of $[0,1]$-valued potential functions G_n on some state space E_n, and some time inhomogeneous Markov chain with elementary transitions

$$\mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right) = M_n(x_{n-1}, dx_n)$$

from E_{n-1} into E_n, with $\mathrm{Law}(X_0) = \eta_0$, is given by the formula

$$\begin{aligned} \mathbb{Q}_n(d(x_0,\ldots,x_n)) &:= \frac{1}{\mathcal{Z}_n}\left\{\prod_{0\le k<n} G_k(x_k)\right\}\mathbb{P}\left((X_0,\ldots,X_n)\in d(x_0,\ldots,x_n)\right) \\ &= \frac{1}{\mathcal{Z}_n}\,\eta_0(dx_0)\,Q_1(x_0,dx_1)\times\ldots\times Q_n(x_{n-1},dx_n) \end{aligned} \tag{2.1}$$

with the integral operators

$$Q_n(x_{n-1},dx_n) = G_{n-1}(x_{n-1})\;M_n(x_{n-1},dx_n)$$

We also denote by η_n the n-th time marginal of $\mathbb{Q}_n$, and $\gamma_n = \mathcal{Z}_n\times\eta_n$. More formally, these distributions are defined for any bounded function f_n by

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f_n) = \mathbb{E}\left(f_n(X_n)\prod_{0\le k<n} G_k(X_k)\right) \tag{2.2}$$

Notice that

$$(2.1) \Longrightarrow \gamma_n = \gamma_{n-1}Q_n \quad \text{and} \quad \eta_n = \Phi_n(\eta_{n-1}) := \Psi_{G_{n-1}}(\eta_{n-1})M_n \tag{2.3}$$

with the Boltzmann-Gibbs transformation $\Psi_{G_{n-1}}$ defined in (7). In physics, the operators Q_n are called Feynman-Kac propagators to underline the fact that

$$\forall 0\le p\le n \qquad \gamma_n = \gamma_p Q_{p,n} \quad \text{with the semigroup} \quad Q_{p,n} = Q_{p+1}Q_{p+1,n}$$

For indicator functions $G_n = 1_A$ and time homogeneous Markov chains on a finite space this model reduces to the one discussed in section 1.5.1. The normalizing constant $\mathcal{Z}_n$ can be described in terms of the measures η_n with the product formula

$$\mathcal{Z}_n = \gamma_n(1) = \mathbb{E}\left(\prod_{0\le k<n} G_k(X_k)\right) = \prod_{0\le p<n}\eta_p(G_p) \tag{2.4}$$

We refer to section 4.2.3 for an elementary proof of this formula.

For absolutely continuous models of the form

$$G_{n-1}(x_{n-1})M_n(x_{n-1}, dx_n) = H_n(x_{n-1}, x_n)\ \lambda_n(dx_n)$$

for some density function H_n w.r.t. some reference measure λ_n we also have the backward formula

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n)\ \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}}(x_q, dx_{q-1})$$

with the collection of backward Markov transitions

$$\mathbb{M}_{n+1,\eta_n}(x_{n+1}, dx_n) \ = \ \frac{\eta_n(dx_n)\ H_{n+1}(x_n, x_{n+1})}{\eta_n\left(H_{n+1}(., x_{n+1})\right)} \tag{2.5}$$

A detailed proof is provided in section 4.2.4.

2.2 Kalman integration models

2.2.1 The Kalman filter


In some instances the path space measures presented in section 2.1 can be computed explicitly. For instance, suppose that X_n is an $\mathbb{R}^p$-valued Markov chain defined by the recursion

$$X_n = A_n\ X_{n-1} + a_n + B_n W_n$$

or some $\mathbb{R}^{d_w}$-valued independent random sequences W_n, independent of X_0, some matrices A_n and B_n with appropriate dimensions and finally some p-dimensional vector a_n. We further assume that W_n centered Gaussian random sequences with covariance matrices R_n^w, and X_0 is a Gaussian random variable in $\mathbb{R}^p$ with a mean and covariance matrix denoted by

$$\widehat{X}_0^- \ = \ \mathbb{E}(X_0) \quad \text{and} \quad \widehat{P}_0^- = \mathbb{E}((X_0 - \mathbb{E}(X_0))\ (X_0 - \mathbb{E}(X_0))')$$

We let y_n, c_n be a given sequence of $\mathbb{R}^q$ vectors, C_n an $\mathbb{R}^{q\times p}$-matrix and $R_n \in \mathbb{R}^{q\times q}$ a symmetric definite positive matrix.

$$G_n(x_n) \propto \exp\{\left(-\frac{1}{2}\left(y_n - [C_n x_n + c_n]'\ R_n^{-1}\ [C_n x_n + c_n]\right)\right)\}$$

In this situation, the n-th time marginals η_n of (2.1) are Gaussian and they coincide with the conditional distributions

$$\eta_n := \text{Law}\left(X_n \mid Y_k = y_k,\ k < n\right) = \mathcal{N}\left(\widehat{X}_n^-, P_n^-\right)$$

and

$$\widehat{\eta}_n := \Psi_{G_n}(\eta_n) := \text{Law}\left(X_n \mid Y_k = y_k,\ k \leq n\right) = \mathcal{N}\left(\widehat{X}_n, P_n\right)$$

of a linear Gaussian filtering problem with the observations sequence

$$Y_n = C_n\ X_n + c_n + V_n\,, \qquad\qquad n \geq 0$$

In the above display, V_n stands for a centered q-dimensional Gaussian random sequence (independent of X) with covariance matrices R_n. The mean vectors $\widehat{X}_n^-, \widehat{X}_n$ and the covariance matrices P_n^-, P_n of

the Gaussian distributions $\mathcal{N}\left(\widehat{X}_n^-, P_n^-\right)$ and $\mathcal{N}\left(\widehat{X}_n, P_n\right)$ are computed using the Kalman recursions provided in section 8.3.1. These prediction-updating equations are given by

$$\begin{aligned}\widehat{X}_n^- &= A_n\,\widehat{X}_{n-1} + a_n \qquad P_n^- = A_n P_{n-1} A_n' + B_n R_n^w B_n' \\ \widehat{X}_n &= \widehat{X}_n^- + P_n^- C_n' \Sigma_n (P_n^-)^{-1}\,\left(Y_n - \left(C_n \widehat{X}_n^- + c_n\right)\right)\end{aligned}$$

with

$$P_n = (I - P_n^- C_n' \Sigma_n (P_n^-)^{-1} C_n) P_n^- \quad \text{and} \quad \Sigma_n(P_n^-) := C_n P_n^- C_n' + R_n$$

For scalar models ($p = 1$) with

$$c_n = a_n = 0 \qquad A_n = B_n = C_n = 1 \qquad R_n = \tau^2 \quad \text{and} \quad R_n^w = \sigma^2$$

the prediction formulae reduce to $\widehat{X}_n^- = \widehat{X}_{n-1}$, $P_n^- = P_{n-1} + \sigma^2$, so that the updating equation takes the simple form

$$\widehat{X}_n = \frac{\tau^2}{\tau^2 + \sigma^2 + P_{n-1}}\,\widehat{X}_{n-1} + \frac{\sigma^2 + P_{n-1}}{\tau^2 + \sigma^2 + P_{n-1}}\,Y_n \quad \text{and} \quad P_n = \frac{\tau^2\left(\sigma^2 + P_{n-1}\right)}{\tau^2 + \sigma^2 + P_{n-1}}$$

Equivalently, we have

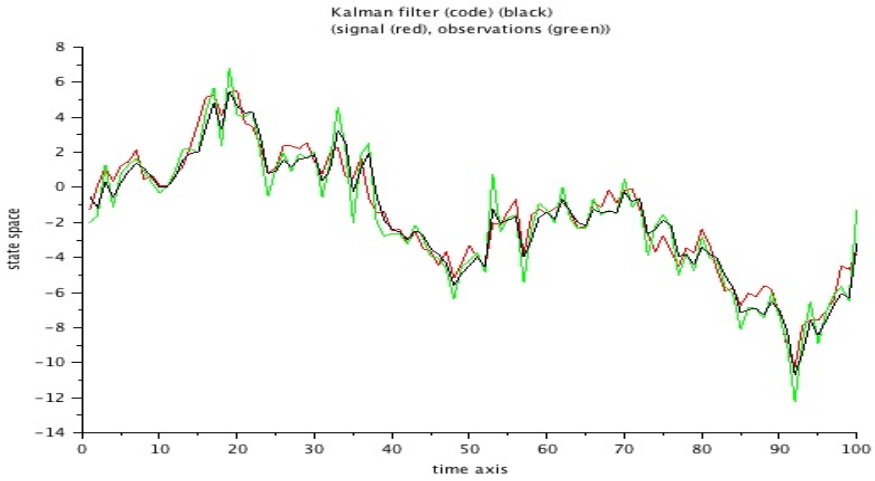
$$\widehat{X}_{n+1}^- = \frac{\tau^2}{\tau^2 + P_n^-}\,\widehat{X}_n^- + \frac{P_n^-}{\tau^2 + P_n^-}\,Y_n \quad \text{and} \quad P_{n+1}^- = \sigma^2 + \frac{\tau^2 P_n^-}{\tau^2 + P_n^-}$$

Notice that in this situation, the one step-optimal predictor takes the form

$$\begin{aligned}\eta_n(f) &\propto \mathbb{E}\left(f(X_n)\,\exp\left(-\frac{1}{2}\sum_{0\le k<n}(y_k - X_k)^2\,\tau^{-2}\right)\right) \quad \text{with} \quad X_n - X_{n-1} = \sigma\,\mathcal{W}_n \\ &\propto \int f(x_n)\,\exp\left(-\frac{1}{2}\,(x_n - \widehat{X}_n^-)^2 (P_n^-)^{-1}\right) dx_n \qquad (2.6)\end{aligned}$$

where $\mathcal{W}_n$ stands for a sequence of i.i.d. centered Gaussian r.v. with unit variance.

The next picture illustrates a realization of a $1d$-Kalman filter associated with $\sigma^2 = 1 = \tau^2$.



2.2.2 The Ensemble Kalman filter

For large dimensional signal processes, such as those arising in data assimilation problems, the covariance matrix P_n^- of the Kalman filter is often too complex to compute. One strategy is to replace this

matrix by an empirical matrix based on a particle interpretation of the Kalman recursions. The resulting stochastic model is defined in terms of an $(\mathbb{R}^p)^N$-valued Markov chain with an updating-prediction transition

$$\xi_n := \left(\xi_n^i\right)_{1\leq i\leq N} \xrightarrow{\text{updating}} \widetilde{\xi}_n := \left(\widetilde{\xi}_n^i\right)_{1\leq i\leq N} \xrightarrow{\text{prediction}} \xi_{n+1} := \left(\xi_{n+1}^i\right)_{1\leq i\leq N}$$

defined for any $1 \leq i \leq N$ by the equations

$$\begin{array}{rcl} \widetilde{\xi}_n^i & = & \xi_n^i + P^-(\xi_n)C_n'\Sigma_n(P^-(\xi_n))^{-1}\ (y_n - C_n\xi_n^i - c_n - V_n^i) \\ \xi_{n+1}^i & = & A_{n+1}\widetilde{\xi}_n^i + a_n + B_{n+1}W_{n+1}^i \end{array}$$

with i.i.d. copies $(V_n^i, W_n^i)_{1\leq i\leq N}$ of (V_n, W_n), and

$$P^-(\xi_n) = \frac{1}{N}\sum_{1\leq i\leq N}\left(\xi_n^i - \frac{1}{N}\sum_{1\leq j\leq N}\xi_n^j\right)\left(\xi_n^i - \frac{1}{N}\sum_{1\leq k\leq N}\xi_n^k\right)'$$

As underlined in [405], in practice the $(p\times p)$-matrix $P^-(\xi_n)$ is never computed or stored. To be more precise, we consider the $q-$column vectors

$$\forall 1\leq i\leq N \qquad \zeta_n^i := C_n\left(\xi_n^i - \frac{1}{N}\sum_{1\leq j\leq N}\xi_n^j\right)$$

These vectors can be evaluated using q-scalar products in $\mathbb{R}^p$ induced by the q-row vectors of the matrix C_n. In this notation, we have

$$\begin{array}{rcl} P^-(\xi_n)C_n' & = & \frac{1}{N}\sum_{1\leq i\leq N}\left(\xi_n^i - \frac{1}{N}\sum_{1\leq j\leq N}\xi_n^j\right)\left(\zeta_n^i\right)' \\ \Sigma_n(P^-(\xi_n)) & = & C_nP^-(\xi_n)C_n' + R_n = \frac{1}{N}\sum_{1\leq i\leq N}\zeta_n^i\left(\zeta_n^i\right)' + R_n \end{array}$$

Therefore, to evaluate $P^-(\xi_n)C_n'$ and $\Sigma_n(P^-(\xi_n))$ only $(N\times q)$ scalar products in $\mathbb{R}^p$ need to be computed.

We refer to section 11.2.3 for a more thorough discussion on these Ensemble Kalman filters. We also refer the reader to section 8.3.2 for a description of the Feynman-Kac measures on path space (2.1) in terms of a backward Gaussian Markov chain model

2.2.3 Interacting Kalman filters

We consider a Markov chain Θ_n evolving in some state spaces Ξ_n, with Markov transitions $K_n(\theta_{n-1}, d\theta_n)$ and an initial condition $\mu_0(d\theta_0)$. We let (Θ_n, X_n, Y_n) be the filtering problem defined as in section 2.2.1 by replacing (A_n, B_n, C_n, R_n) and (a_n, c_n) by $(A_n(\Theta_n), B_n(\Theta_n), C_n(\Theta_n), R_n(\Theta_n))$ and $(a_n(\Theta_n), c_n(\Theta_n))$. Given a realization of the historical chain $\boldsymbol{\Theta_n} = (\Theta_0, \ldots, \Theta_n)$, the one step optimal predictor is given by

$$\eta_{\boldsymbol{\Theta_n},n} := \mathrm{Law}\left(X_n \mid \boldsymbol{\Theta_n},\ Y_k,\ k<n\right) = \mathcal{N}\left(\widehat{X}_n^{\boldsymbol{\Theta_n},-}, P_n^{\boldsymbol{\Theta_n},-}\right)$$

with the mean-covariance parameters $\left(\widehat{X}_n^{\boldsymbol{\Theta_n},-}, P_n^{\boldsymbol{\Theta_n},-}\right)$ computed using the Kalman recursions associated with the parameters $(A_n(\Theta_n), B_n(\Theta_n), C_n(\Theta_n), R_n(\Theta_n))$ and $(a_n(\Theta_n), c_n(\Theta_n))$. By construction, the triplet

$$\widetilde{X}_n := \left(\Theta_n, \widehat{X}_n^{\boldsymbol{\Theta_n},-}, P_n^{\boldsymbol{\Theta_n},-}\right)$$

is a Markov chain. In addition, using Bayesian notation we have

$$\mathbb{P}(Y_n \in dy_n \mid \mathbf{\Theta_n},\ Y_k,\ k<n) = \int\ p(Y_n \mid x_n,\ \Theta_n)\ \eta_{\mathbf{\Theta_n},n}(dx_n)$$

so that

$$\mathrm{Law}(Y_n \mid \mathbf{\Theta_n},\ Y_k,\ k<n) = \mathcal{N}\left(C_n(\Theta_n)\widehat{X}_n^{\mathbf{\Theta_n},-} + c_n(\Theta_n),\ C_n(\Theta_n)P_n^{\mathbf{\Theta_n},-}C_n'(\Theta_n) + R_n(\Theta_n)\right) \quad (2.7)$$

We fix the observation sequence $Y_n = y_n$, with $n \geq 0$, and we set

$$\widetilde{G}_n(\widetilde{X}_n) := \widetilde{g}_n(y_n, \widetilde{X}_n)$$

where $y_n \mapsto \widetilde{g}_n(y_n, \widetilde{X}_n)$ stands for the density of the Gaussian distribution (2.7).

We sample N independent copies $(\zeta_0^i)_{1\leq i\leq N}$ of $\widetilde{X}_0$, and we compute the likelihood weights $\widetilde{G}_0(\zeta_0^i)$, with $1 \leq i \leq N$. For each $1 \leq i \leq N$, with a probability $\widetilde{G}_0(\zeta_0^i)/\max_{1\leq j\leq N}\widetilde{G}_0(\zeta_0^j)$ we set $\widehat{\zeta}_0^i := \zeta_0^i$. Otherwise, set $\widehat{\zeta}_0^i := \widetilde{\zeta}_0^i$, where $\widetilde{\zeta}_0^i$ stands for a random sample from the weighted distribution

$$\sum_{1\leq j\leq N} \frac{\widetilde{G}_0(\zeta_0^j)}{\sum_{1\leq k\leq N}\widetilde{G}_0(\zeta_0^k)}\ \delta_{\zeta_0^j}$$

At the end of this updating step, move independently each $\widehat{\zeta}_0^i$ to a new state ζ_1^i according to the transition of the states $\widetilde{X}_0 \rightsquigarrow \widetilde{X}_1$ starting at $\widehat{\zeta}_0^i$, with $1 \leq i \leq N$.

Compute the likelihood weights $\widetilde{G}_1(\zeta_1^i)$, with $1 \leq i \leq N$. For each $1 \leq i \leq N$, with a probability $\widetilde{G}_1(\zeta_1^i)/\max_{1\leq j\leq N}\widetilde{G}_1(\zeta_1^j)$ we set $\widehat{\zeta}_1^i := \zeta_1^i$. Otherwise, set $\widehat{\zeta}_1^i := \widetilde{\zeta}_1^i$, where $\widetilde{\zeta}_1^i$ stands for a random sample from the weighted distribution

$$\sum_{1\leq j\leq N} \frac{\widetilde{G}_1(\zeta_1^j)}{\sum_{1\leq k\leq N}\widetilde{G}_1(\zeta_1^k)}\ \delta_{\zeta_1^j}$$

At the end of this updating step, move independently each $\widehat{\zeta}_1^i$ to a new state ζ_2^i according to the transition of the states $\widetilde{X}_1 \rightsquigarrow \widetilde{X}_2$ starting at $\widehat{\zeta}_1^i$, with $1 \leq i \leq N$; and so on.

This branching type interacting particle system algorithm can be interpreted as a sequence of N-interacting Kalman filters. Tracing back in time the ancestral lines

$$(\Theta_{0,n}^i, \Theta_{1,n}^i, \ldots, \Theta_{n,n}^i)$$

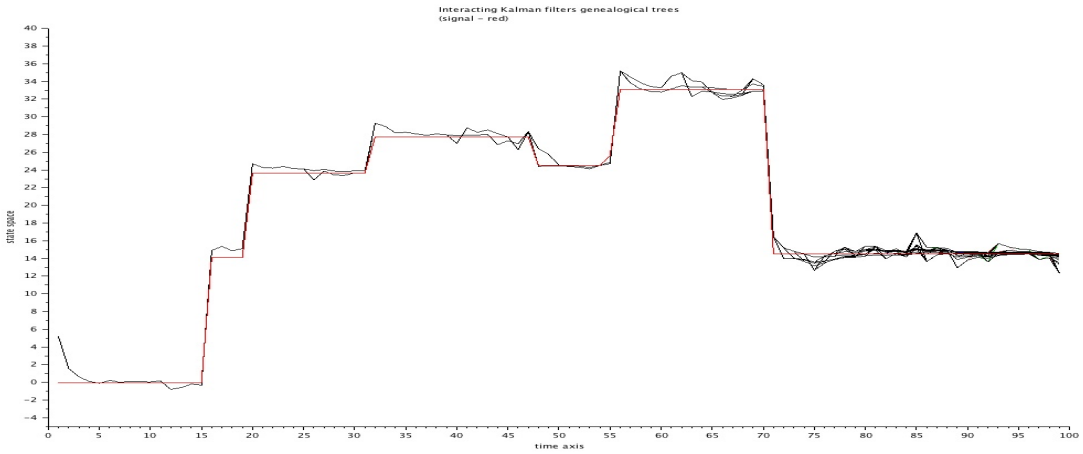
of the Θ-components of the particles ζ_n^i we obtain the genealogical tree based approximations

$$\frac{1}{N}\sum_{1\leq i\leq N}\delta_{(\Theta_{0,n}^i,\Theta_{1,n}^i,\ldots,\Theta_{n,n}^i)}\ \simeq_{N\uparrow\infty}\ \mathrm{Law}((\Theta_0,\ldots,\Theta_n) \mid Y_k = y_k,\ k<n)$$

In addition, if $\eta_{\mathbf{\Theta_n},n}^{(N,i)}$ stands for the Gaussian distributions associated with the $(\widehat{X}_n^{\mathbf{\Theta_n},-}, P_n^{\mathbf{\Theta_n},-})$-mean-covariance components of the particles ζ_n^i, then we have

$$\frac{1}{N}\sum_{1\leq i\leq N}\eta_{\mathbf{\Theta_n},n}^{(N,i)}\ \simeq_{N\uparrow\infty}\ \mathrm{Law}(X_n \mid Y_k = y_k,\ k<n)$$

For non Gaussian perturbations (W_n, V_n), the extended Kalman integration is performed by replacing these variables by Gaussian perturbations with the same variance. The next picture illustrates a realization of the genealogical tree associated with $N = 100$ interacting Kalman filters associated Bernoulli jump type random variables $B(\Theta_n) = \Theta_n$ with rate 10%, and uniform jump amplitudes W_n on $[-20, 20]$, with unit sensor perturbation noise.



These interacting Kalman filters belongs to the class of quenched and annealed Feynman-Kac models discussed in section 10.6.2, and section 11.2.5.

An alternative particle approximation based on genetic type population models is provided in section 2.3.2.

2.3 Evolutionary particle models

The may difficulty to design a Monte Carlo approximation of Feynman-Kac measures comes from the degeneracy of the weight product in (2.1). For instance for indicator functions $G_n = 1_A$ most of the importance sampling strategy will fail to remains in the set A for large time horizons.

2.3.1 Branching processes

The Feynman-Kac measures η_n can be interpreted as the first moment of the occupation measures

$$\mathcal{X}_n = \sum_{i=1}^{p_n} \delta_{\zeta_n^i}$$

of a spatial branching process $(\zeta_n^i)_{1\leq i\leq p_n} \in S = \cup_{p\geq 0} E^p$. We let $(g_n^i(x))_{i\geq 1, x\in E, n\geq 0}$ be a collection of integer number-valued random variables and we set

$$G_n(x) := \mathbb{E}(g_n^i(x))$$

The branching process is defined as follows. We start at some point x_0 with a single particle, that is $p_0 = 1$ and $\zeta_0 = \zeta_0^1 = x_0 \in E^{p_0} = E$.

At each time $n \geq 0$, every individuals ζ_n^i, with $1 \leq i \leq p_n$, branches into $g_n^i(\zeta_n^i)$ offsprings. At the end of the branching transition, we have

$$\widehat{p}_n = \sum_{1\leq i\leq p_n} g_n^i(\zeta_n^i) \quad \text{individuals} \quad \widehat{\zeta}_n = (\widehat{\zeta}_n^1, \ldots, \widehat{\zeta}_n^{\widehat{p}_n}) \in E^{\widehat{p}_n}$$

Each of these individuals $\widehat{\zeta}_n^i$ explores randomly the state space E, according to the transition M_n. At the end of this mutation step, we have a population of $p_n = \widehat{p}_n$ particles $\zeta_n^i \in E$ with distribution $M_n(\widehat{\zeta}_n^i, .)$, $i = 1, \ldots, p_n$. By construction, we have

$$\mathbb{E}(\mathcal{X}_{n+1}(f) \mid \zeta_n) = \mathcal{X}_n(G_n M_{n+1}(f))$$

and therefore

$$\mathbb{E}(\mathcal{X}_n(f)) = \mathbb{E}_{x_0}\left(f(X_n)\prod_{0\leq k<n} G_k(X_k)\right) := \gamma_n(f)$$

In the above display, X_n stands for the Markov chain on E with Markov transitions M_n. In this interpretation, the mean number of individuals in the current population is given by $\mathbb{E}(\mathcal{X}_n(1)) = \gamma_n(1)$. For a more detailed discussion on spatial branching models and their applications in multiple-object filtering problems, we refer the reader to section 12.1.1

2.3.2 A fixed population size branching process

The mutation-selection transitions of the particle interpretation of a general Feynman-Kac model of the form (2.1) is defined as follows:

- During the selection stage, we quote $p_n^N := \frac{1}{N}\sum_{1\leq i\leq N} G_n(\xi_n^i)$ the empirical average of the particles potential values $G_n(\xi_n^i)$. When $p_n^N \neq 0$ each particle ξ_n^i is accepted with a probability $G_n(\xi_n^i)$ and we set $\widehat{\xi}_n^i = \xi_n^i$. Rejected particles are resampled by choosing randomly a state according to the weighted discrete distribution

$$\sum_{1\leq i\leq N} \frac{G_n(\xi_n^i)}{\sum_{1\leq j\leq N} G_n(\xi_n^j)}\ \delta_{\xi_n^i} \tag{2.8}$$

  When $p_n^N = 0$, the algorithm need to be restarted.

- During the mutation stage, each particles $\widehat{\xi}_n^i$ moves independently according to the Markov transitions M_{n+1}; that is, we sample N independent r.v. ξ_{n+1}^i with distribution $M_{n+1}\left(\widehat{\xi}_n^i, dx_{n+1}\right)$.

In this situation, the distribution $\mathbb{Q}_n$ can be approximated by the empirical measures of the ancestral line of the individuals, that is

$$\frac{1}{N}\sum_{1\leq j\leq N} \delta_{(\xi_{0,n}^j,\ldots,\xi_{n,n}^j)} \ \simeq_{N\uparrow\infty}\ \mathbb{Q}_n$$

We also have the unbiasedness particle estimates

$$\mathcal{Z}_n^N := \prod_{0\leq k<n}\frac{1}{N}\sum_{1\leq i\leq N} G_k(\xi_k^i) \ \simeq_{N\uparrow\infty}\ \mathcal{Z}_n$$

These genetic type interacting particle systems have been used with success in a variety of application domains as heuristic like Monte Carlo schemes since the end of the 1940s. We quote the pioneering articles by T.E. Harris and H. Kahn [338], published in 1951, and the one by Enrico Fermi and R.D. Richtmyer in 1948 on Resampled type Quantum Monte Carlo methodologies. To the best of our knowledge, the first rigorous mathematical foundations of these models have been published in 1996 in [160] (seel also [161]).

Depending on their application domains the genetic type selection-mutation transitions discussed above are also known under different guises, with a variety of different names and terminologies. For instance the r.v. ξ_n^i are called *samples, particles, individuals, or replica*. To guide the reader in these interdisciplinary literature, in the following table we have tried to summarize some more or less equivalent formulations of the two step transitions of the algorithm discussed above.

Sequential Monte Carlo	Sampling	Resampling
Particle Filters	Prediction	Updating
Data assimilation	Forecasting	Analysis
Genetic Algorithms	Mutation	Selection
Evolutionary Population	Exploration	Branching-selection
Diffusion Monte Carlo	Free evolutions	Absorption
Quantum Monte Carlo	Walkers motions	Reconfiguration
Sampling Algorithms	Transition proposals	Accept-reject-recycle

The selection transition in the r.h.s. column is also termed: bootstrapping, spawning, cloning, pruning, replenish, multi-level splitting, enrichment, go with the winner, quantum teleportation,...

For a more detailed discussion on particle Feynman-Kac we refer to section 3.3.1, section 9.1.4, as well as to chapter 10.

2.3.3 Island particle models

We consider the particle model discussed in section 2.3.2. In section 10.1, for any bounded function f_n on E_n we prove that

$$\mathbb{E}\left(\eta_n^N(f_n)\prod_{0\leq k<n}\eta_k^N(G_k)\right)=\mathbb{E}\left(f_n(X_n)\prod_{0\leq k<n}G_k(X_k)\right) \tag{2.9}$$

with

$$\eta_n^N:=\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i}\Longrightarrow\eta_n^N(f_n)=\int f(x_n)\,\eta_n^N(dx_n)=\frac{1}{N}\sum_{1\leq i\leq N}f_n(\xi_n^i)$$

If we set

$$\chi_n:=\xi_n\qquad\mathcal{F}_n(\chi_n):=\eta_n^N(f_n)\quad\text{and}\quad\mathcal{G}_n(\chi_n):=\eta_n^N(G_n)$$

then (2.9) takes the form

$$\mathbb{E}\left(\mathcal{F}_n(\chi_n)\prod_{0\leq k<n}\mathcal{G}_k(\chi_k)\right)=\mathbb{E}\left(f_n(X_n)\prod_{0\leq k<n}G_k(X_k)\right)$$

The l.h.s. is a Feynman-Kac formula with reference Markov chain χ_n and potential functions $\mathcal{G}_n$. The particle approximation of these models are in terms of an interacting island particle model. During the mutation stage, the islands evolve independently according the Markov transitions of the particle model $\chi_n=\xi_n$. During the selection stage, the island are accepted or rejected depending on the empirical averages $\mathcal{G}_k(\chi_k)$ of the individuals within each island. These many-body Feynman-Kac models are discussed in section 10.5.

2.3.4 Backward particle models

We further assume that the Markov transitions M_n are absolutely continuous with respect to some measures λ_n on E_n, and for any $(x_{n-1},x_n)\in(E_{n-1}\times E_n)$ we have

$$G_{n-1}(x_{n-1})\;M_n(x_{n-1},dx_n)=H_n(x_{n-1},x_n)\;\lambda_n(dx_n) \tag{2.10}$$

for some density function H_n. In this situation, we also have that

$$\mathbb{Q}_n^N(d(x_0,\ldots,x_n))=\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i}(dx_n)\prod_{q=1}^{n}\mathbb{M}_q^N(x_q,dx_{q-1})\;\simeq_{N\uparrow\infty}\;\mathbb{Q}_n(d(x_0,\ldots,x_n)) \tag{2.11}$$

with the random Markov transitions

$$\mathbb{M}_q^N(x_q, dx_{q-1}) = \sum_{1\leq i\leq N} \frac{H_q(\xi_{q-1}^i, x_q)}{\sum_{1\leq j\leq N} H_q(\xi_{q-1}^j, x_q)}\ \delta_{\xi_{q-1}^i}(dx_{q-1})$$

By construction, $\mathbb{Q}_n^N$ can be interpreted as the distribution of a Markov chain on $\{1,\ldots,N\}$ evolving backward in time, starting with an uniform distribution at time n, with elementary transitions from time q to time $q-1$ given by the (random) stochastic matrix

$$\begin{pmatrix} \frac{H_q(\xi_{q-1}^1,\xi_q^1)}{\sum_{1\leq j\leq N} H_q(\xi_{q-1}^j,\xi_q^1)} & \cdots & \frac{H_q(\xi_{q-1}^N,\xi_q^1)}{\sum_{1\leq j\leq N} H_q(\xi_{q-1}^j,\xi_q^1)} \\ \vdots & \vdots & \vdots \\ \frac{H_q(\xi_{q-1}^1,\xi_q^N)}{\sum_{1\leq j\leq N} H_q(\xi_{q-1}^j,\xi_q^N)} & \cdots & \frac{H_q(\xi_{q-1}^N,\xi_q^N)}{\sum_{1\leq j\leq N} H_q(\xi_{q-1}^j,\xi_q^1)} \end{pmatrix}$$

In section 10.5.3 (see corollary 10.5.6) we also prove that

$$\mathbb{P}\left(\left(\xi_{0,n}^j,\ldots,\xi_{n,n}^j\right)\in d(x_0,\ldots,x_n)\mid \xi_0,\ldots,\xi_{n-1}\right) = \Phi_n(\eta_{n-1}^N)(dx_n)\ \prod_{q=1}^n \mathbb{M}_q^N(x_q,dx_{q-1})$$

for any $1\leq j\leq N$ with

$$\Phi_n(\eta_{n-1}^N)(dx_n) = \sum_{1\leq i\leq N} \frac{G_{n-1}(\xi_{n-1,n-1}^i)}{\sum_{1\leq j\leq N} G_{n-1}(\xi_{n-1,n-1}^j)}\ M_n(\xi_{n-1,n-1}^i, dx_n)$$

These backward Feynman-Kac integration models are developed in section 10.5.4, section 4.2.4 and section 10.3 (see also section 8.3.2 for a derivation of these models in the context of linear-Gaussian filtering problems). A more detailed discussion on these Feynman-Kac particle models is provided in chapter 10.

2.3.5 Particle Markov chain Monte Carlo models

A couple of Markov chain Monte Carlo samplers with target measure $\mathbb{Q}_n$ can be underlined:

- As in (1.9), starting from a given trajectory $\mathbb{X}_0 = (\xi_{k,n}^1)_{0\leq k\leq n}$, we sample a new genetic type model $\xi_n = \left(\xi_n^i\right)_{1\leq i\leq N}$ defined as above with a frozen trajectory $\mathbb{X}_0$. More precisely, the first particle at time $k(\leq n)$ coincides with the k-th ancestor $\xi_{k,n}^1$; and the acceptance-rejection/selection transition defined in (2.8) only concern the $(N-1)$ remaining particles. Then, we choose randomly a genealogical line $\mathbb{X}_1$ or we sample backward in time a ancestral line $\mathbb{X}_1$ according to the backward particle Markov chain model defined in (2.11). The Markov transition $\mathbb{X}_0 \rightsquigarrow \mathbb{X}_1$ on path space is reversible w.r.t. $\mathbb{Q}_n$. Thus, iterating this procedure, we define a Markov chain on path space with invariant measure $\mathbb{Q}_n$.

- We start with a realization of the particle model $\mathcal{X}_0 := \xi := (\xi_k)_{0\leq k\leq n}$. Then we sample a new independent particle model $\overline{\xi} := (\overline{\xi}_k)_{0\leq k\leq n}$ and we accept $\overline{\xi}$ with probability

$$a(\xi,\overline{\xi}) := \min\left(1, \prod_{0\leq k<n} \frac{\frac{1}{N}\sum_{1\leq i\leq N} G_k(\overline{\xi}_k^i)}{\frac{1}{N}\sum_{1\leq i\leq N} G_k(\xi_k^i)}\right)$$

  In this case we set $\mathcal{X}_1 = \overline{\xi}$; otherwise $\mathcal{X}_1 = \mathcal{X}_0$. The Markov transition $\mathcal{X}_0 \rightsquigarrow \mathcal{X}_1$ on path space is reversible w.r.t. some measure on the product space $(E_0\times\ldots\times E_n)^N$ with $(E_0\times\ldots\times E_n)$-marginals equal to $\mathbb{Q}_n$. Thus, iterating this procedure, we define a Markov chain on path space with invariant measure $\mathbb{Q}_n$.

For a detailed discussion on these particle Markov chain Monte Carlo methodologies, we refer the reader to section 10.5.2.

2.4 Twisted models and h-processes

2.4.1 Spectral decompositions

We consider homogeneous models $(E_n, G_n, M_n) = (E, G, M)$ equipped with a reversible transition M w.r.t. some probability measure μ. Under some rather weak regularity conditions, the integral operator $Q(x, dy) = G(x)M(x, dy)$ on $\mathbb{L}_2(\mu)$ has a $h(x)$ a positive eigenvector associated with the maximal eigenvalue λ; that is, we have that

$$Q(h)(x) = G(x)M(h)(x) = \lambda h(x) \implies G(x) = \lambda\, h(x)/M(h)(x)$$

To simplify the presentation and avoid some unnecessary technical discussion on the integrability of the potential function w.r.t. reference measure μ, we further assume that $\epsilon \leq G \leq \epsilon^{-1}$, for some $\epsilon > 0$, and $M(x, dy) = m(x, y)\mu(dy)$ for some density function $m \in \mathbb{L}_2(\mu \otimes \mu)$. In this situation, Q is a compact self-adjoint operator on $\mathbb{L}_2(\mu) = \mathbb{L}_2(\Psi_{G^{-1}}(\mu))$ (with $G^{-1} = 1/G$). The $\Psi_{G^{-1}}(\mu)$-reversibility of Q comes from the fact that for any functions $f_1, f_2 \in \mathbb{L}_2(\mu)$ we have

$$\Psi_{G^{-1}}(\mu)\,(f_1 Q(f_2)) \propto \mu(f_1 M(f_2)) = \mu(M(f_1)\ f_2) \propto \Psi_{G^{-1}}(\mu)\,(Q(f_1)\ f_2)$$

The spectral theorem for compact and self adjoint and positive operators allows to rewrite any power Q^n in terms of a countable (by the compactness property) orthonormal basis $(\varphi_i)_{i\geq 0} \in \mathbb{L}_2(\Psi_{G^{-1}}(\mu))$ of eigenfunctions associated with a non increasing sequence of positive eigenvalues $(\lambda_i)_{i\geq 0}$ of Q

$$Q^n(x, dy) = \sum_{i\geq 0} \lambda_i^n\ \varphi_i(x)\varphi_i(y)\ \Psi_{G^{-1}}(\mu)(dy) \quad \text{with} \quad (\lambda_0, \varphi_0) = (\lambda, h) \tag{2.12}$$

In this situation, in section 4.3.2 we prove that

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \frac{1}{\mathbb{E}(h^{-1}(X_n^h))}\ h^{-1}(x_n)\ \ \mathbb{P}_n^h(d(x_0, \ldots, x_n)) \tag{2.13}$$

where $\mathbb{P}_n^h = \text{Law}\left(X_0^h, \ldots, X_n^h\right)$ stands for the distribution of the random path of a Markov chain X_n^h with initial distribution $\eta_0^h = \Psi_h(\eta_0)$ and Markov transitions

$$\mathbb{P}\left(X_n^h \in dx_n \mid X_{n-1}^h = x_{n-1}\right) = M^h(x_{n-1}, dx_n) := \frac{M(x, dy)h(y)}{M(h)(x)}$$

In this situation, the sampling of (2.1) reduces to that of sampling the h-process X_n^h. It is also readily check that M^h is reversible w.r.t. $\Psi_{hM(h)}(\mu)$

$$\Psi_{hM(h)}(\mu)\left(f_1 M^h(f_2)\right) \propto \mu((hf_1)M(hf_2)) = \mu(M(hf_1)\ (hf_2)) \propto \Psi_{hM(h)}(\mu)\left(M^h(f_1)\ f_2\right)$$

Last but not least, we observe that

$$\begin{aligned} \eta_n(f) \quad &= \quad \frac{\mathbb{E}\left(h^{-1}(X_n^h)f(X_n^h)\right)}{\mathbb{E}\left(h^{-1}(X_n^h)\right)} \\ &\simeq_{n\uparrow\infty} \quad \frac{\Psi_{hM(h)}(\mu)(h^{-1}f)}{\Psi_{hM(h)}(\mu)(h^{-1})} = \frac{\mu(M(h)f)}{\mu(M(h))} = \Psi_{M(h)}(\mu)(f) := \eta_\infty(f) \end{aligned}$$

Using the fact that M is μ-reversible, we also find that

$$\mu(M(h)f) = \mu(hM(f))$$

$$\Rightarrow\ \eta_\infty(f) = \Psi_{M(h)}(\mu)(f) = \Psi_h(\mu)M(f) = \frac{\Psi_{G^{-1}}(\mu)(hf)}{\Psi_{G^{-1}}(\mu)(h)} \qquad (\Leftarrow h/G \propto M(h))$$

We also notice that

$$\Psi_G(\eta_\infty) = \Psi_h(\mu)$$

Finally using (2.12) we have

$$\lambda_0 > \lambda_1 \Longrightarrow \eta_n(f) = \eta_\infty(f) + \mathrm{O}\left((\lambda_1/\lambda_0)^n\right)$$

and by (2.4) we find that

$$\frac{1}{n}\log \mathcal{Z}_n = \frac{1}{n}\log \gamma_n(1) = \frac{1}{n}\sum_{0\le p<n} \log \eta_n(G) = \log \eta_\infty(G) + \mathrm{O}\left(1/n\right) = \log \lambda + \mathrm{O}\left(1/n\right)$$

The r.h.s. formula comes from the fact that

$$\eta_\infty(G) = \Psi_{M(h)}(\mu)(G) = \frac{\mu(GM(h))}{\mu(M(h))} = \lambda \qquad (\Leftarrow G\ M(h) = \lambda h)$$

The measure η_∞ can be approximated using the genetic type particle models ξ_n defined in section 2.3.2. More formally, we have

$$\eta_n^N := \frac{1}{N}\sum_{1\le i\le N} \delta_{\xi_n^i} \ \simeq_{N\uparrow\infty}\ \eta_n \simeq_{n\uparrow\infty} \eta_\infty$$

as well as the empirical spatio-temporal averages

$$\frac{1}{n}\sum_{0\le k<n} \eta_k^N \ \simeq_{N\uparrow}\ \frac{1}{n}\sum_{0\le k<n} \eta_k \ \simeq_{n\uparrow\infty}\ \eta_\infty$$

2.4.2 Guiding and pilot functions

Unfortunately, most of the time in the function h is unknown. One strategy is to use a judiciously chosen trial function $h_{\mathcal{T}}$. In this case, we have

$$\mathbb{Q}(d(x_0,\ldots,x_n)) \propto \left\{\prod_{0\le k<n} G(x_k) h_{\mathcal{T}}^{-1}(x_k) M(h_{\mathcal{T}})(x_k)\right\}\ h_{\mathcal{T}}^{-1}(x_n)\ \mathbb{P}_n^{h_{\mathcal{T}}}(d(x_0,\ldots,x_n)) \tag{2.14}$$

For instance, choosing $h_{\mathcal{T}} = G$ we have

$$\mathbb{Q}(d(x_0,\ldots,x_n))\ G(x_n) \propto \left\{\prod_{0\le k<n} \widehat{G}(x_k)\right\}\ \widehat{\mathbb{P}}_n(d(x_0,\ldots,x_n))$$

with the potential function $\widehat{G} = M(G)$ the distribution $\widehat{\mathbb{P}}_n = \mathrm{Law}(\widehat{X}_0,\ldots,\widehat{X}_n)$ of a Markov chain with transition probabilities

$$\widehat{M}(x_{n-1}, dx_n) = \frac{M(x_{n-1}, dx_n)G(x_n)}{M(G)(x_{n-1})} \tag{2.15}$$

On natural way to evaluate the functions $M(h_{\mathcal{T}})(x)$ and sample the Markov chain transitions $X_{k-1}^{h_{\mathcal{T}}} = x \rightsquigarrow X_k^{h_{\mathcal{T}}}$ consists in replacing the Markov transitions M by the empirical transitions associated with N independent random variables $X_k^{h_{\mathcal{T}},i}(x)$ with common law $M^{h_{\mathcal{T}}}(x, dy)$; that is, we have that

$$M^{h_{\mathcal{T}},N}(x, dy) := \frac{1}{N}\sum_{1\le i\le N} \delta_{X_k^{h_{\mathcal{T}},i}(x)}(dy)\ \simeq_{N\uparrow\infty}\ M^{h_{\mathcal{T}}}(x, dy)$$

For a more thorough discussion on importance sampling schemes and change of probability measures we refer the reader to section 4.1.1.

2.5 Some illustrations

We illustrate the rather abstract model (2.1) with a series of examples.

2.5.1 Self avoiding walks

The first one concerns the simulation of a self-avoiding walk X'_n on $\mathbb{Z}^2$ starting at the origin. In this situation, the Markov chain X_n and the potential functions G_n are given by

$$X_n = \left(X'_0, \ldots, X'_n\right) \in E_n := (\mathbb{Z}^2)^{n+1} \quad \text{and} \quad G_n(X_n) = 1_{X'_n \notin \{X'_0, \ldots, X'_{n-1}\}} \tag{2.16}$$

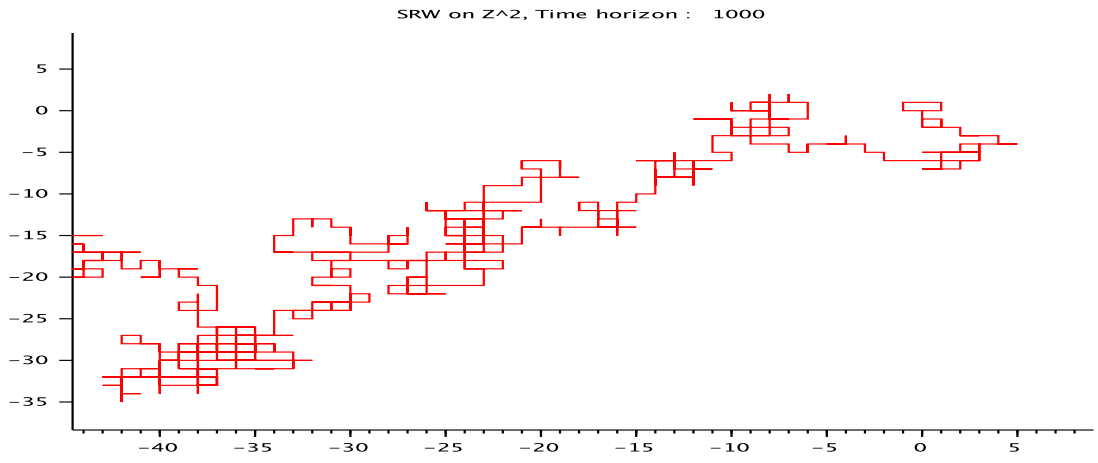
In this case, we notice that

$$\mathbb{Q}_n := \text{Law}\left((X'_0, \ldots, X'_n) \mid \forall 0 \leq k \neq l < n \; X'_k \neq X'_l\right)$$

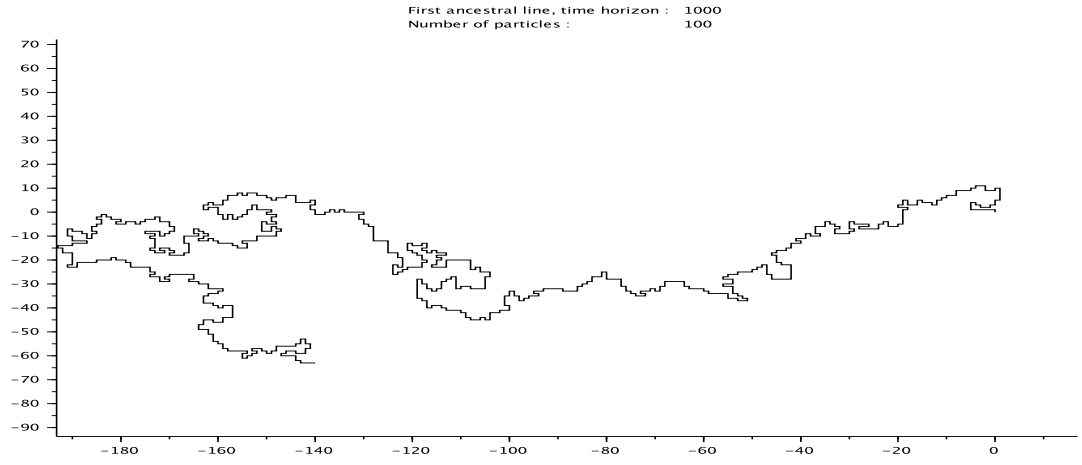
and $\mathcal{Z}_{n+1} = 4^{-n}$ Card$(\mathcal{A}_n)$ with

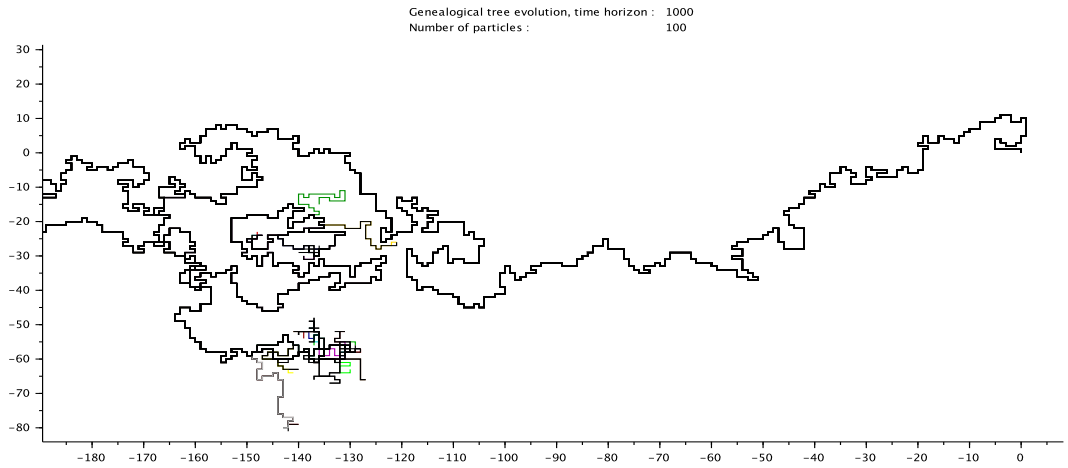
$$\mathcal{A}_n := \left\{(x_0, \ldots, x_n) \in (\mathbb{Z}^2)^{n+1} \mid \forall 0 \leq k \neq l \leq n \quad |x_k - x_{k-1}| = 1 \quad x_k \neq x_l\right\}$$

The following picture illustrates a sample of the SRW on $\mathbb{Z}^2$ on the time horizon $[0, 1000]$



The following pictures illustrates an ancestral line and the genealogical tree based particle model with $N = 100$ non intersection on $\mathbb{Z}^2$ on the time horizon $[0, 1000]$.





These particle models have been simulated using the free evolution of the random walk and the selection indicator potential functions (2.16). Alternatively, we can choose any stretching type change of random walk motion. For instance, the random walk with transition (2.15) (often termed the myopic random walk) evolve using local transitions that avoid the states visited in the past. In this case, the selection potential function $\widehat{G}$ evaluates the chances to avoid these historical states in the next local transition.

Using the fact that $\mathrm{Card}(\mathcal{A}_{p+q}) \leq \mathrm{Card}(\mathcal{A}_p) \times \mathrm{Card}(\mathcal{A}_q)$ and $2^n \leq \mathrm{Card}(\mathcal{A}_n) \leq 4 \times 3^n$, using sub-additivity arguments we find that

$$c := \lim_{n \to \infty} \mathrm{Card}(\mathcal{A}_n)^{1/n} \in [2, 3]$$

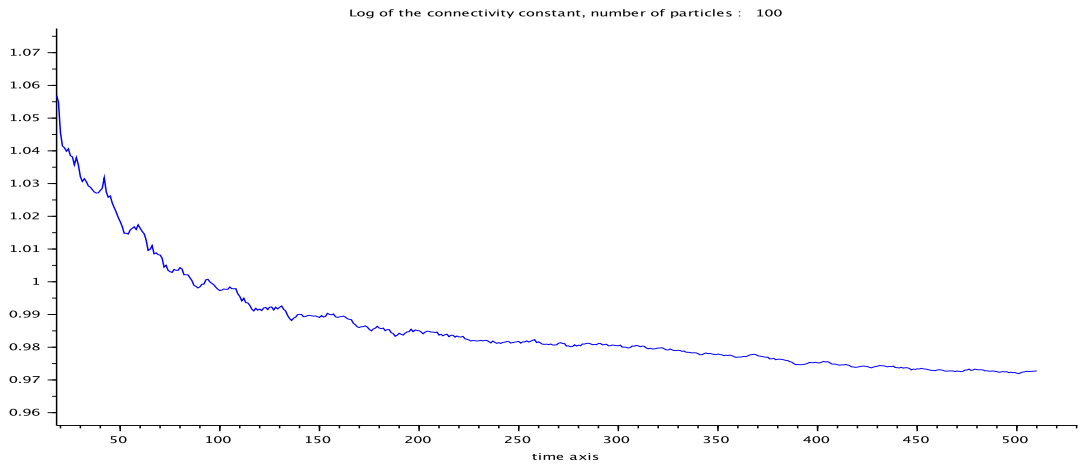
To estimate the co-called connectivity constant c (a.k.a. the critical fugacy), we use the fact that

$$\eta_n(G_n) \quad = \quad \mathbb{P}\left(X' \notin \{X'_0, \ldots, X'_{n-1}\} \mid \forall 0 \leq p < q < n \; X'_p \neq X'_q\right) = \mathrm{Card}(\mathcal{A}_n) / \left(4 \; \mathrm{Card}(\mathcal{A}_{n-1})\right)$$

and therefore

$$\frac{1}{n} \log \gamma_n(G_n) = \frac{1}{n} \sum_{0 \leq k \leq n} \log \eta_k(G_k) = \log\left(c/4\right)$$

The following picture provides an estimate of the connectivity constant with $N = 100$ particles.



Self avoiding random walks (*abbreviated SAW*) are used in physics to model the evolution of linear/directed polymers. These polymers represent the formation of long molecules consisting of monomers linked together in a given chemical solvent. The location of the monomers is encoded in the random walk evolution, the time horizon represents the length of the molecule. These models can be extended with little extra work to analyze polymer models in a confined geometry. For instance, given some subsets $A_n \in \mathbb{Z}^2$ the Feynman-Kac models (2.1) associated with the historical process (2.16) and the potential functions

$$G_n(X_n) = 1_{X'_n \notin \{X'_0,\ldots,X'_{n-1}\}} \times 1_{A_n}(X'_n)$$

are given by

$$\mathbb{Q}_n := \mathrm{Law}\left((X'_0,\ldots,X'_n) \mid \forall 0 \leq k \neq l < n \quad X'_k \neq X'_l \text{ and } (X'_0,\ldots,X'_{n-1}) \in (A_0 \times \ldots \times A_{n-1})\right)$$

and

$$\mathcal{Z}_n = \mathbb{P}\left(\ \forall 0 \leq k \neq l < n \quad X'_k \neq X'_l \text{ and } (X'_0,\ldots,X'_{n-1}) \in (A_0 \times \ldots \times A_{n-1})\right)$$

2.5.2 Particle filters

The second one is related to nonlinear filtering models. We consider a couple signal-observation model $(X_n, Y_n) \in \mathbb{R}^2$ satisfying the following recursion

$$\left\{ \begin{array}{rcl} X_n & = & a_n(X_{n-1}, W_n) \\ Y_n & = & b_n(X_n) + V_n \end{array} \right.$$

In the above display, W_n and V_n stands for a sequence of i.i.d.centered Gaussian r.v. with unit variance. The functions a_n and b_n stands for some regular functions. We fix an observation sequence $Y = y$, and for any $n \geq 0$ we set

$$G_n(x) := \frac{1}{\sqrt{2\pi}} \ \exp\left(-\frac{1}{2}(y_n - h_n(x))^2\right)$$

In this situation, using Bayes' rule we prove that the Feynman-Kac measure (2.1) coincides with the posterior distribution of the trajectories of the signal $(X_0,\ldots,X_n)$ given the observation sequence $Y_k = y_k$, up to time n; that is, we have that

$$\mathbb{Q}_n = \mathrm{Law}\left((X_0,\ldots,X_n) \mid (Y_0,\ldots,Y_{n-1}) = (y_0,\ldots,y_{n-1})\right)$$

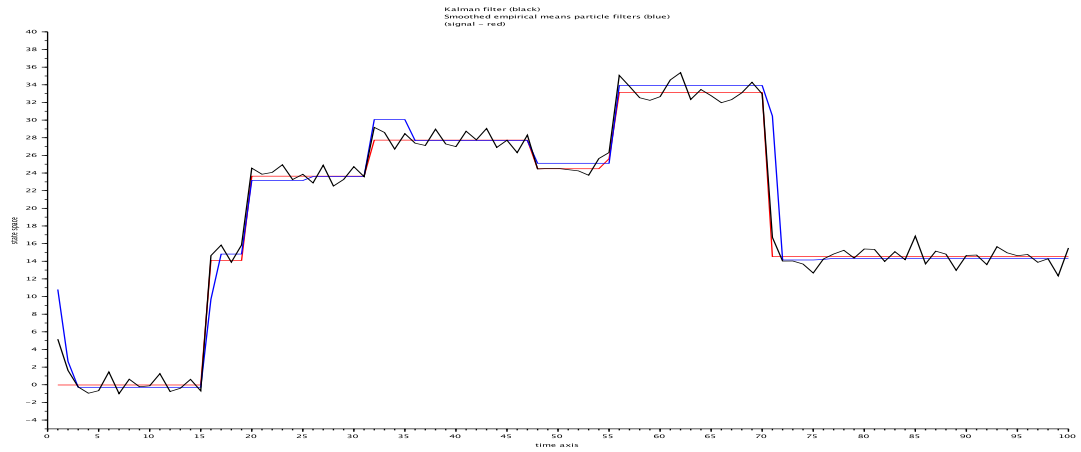
In addition, the density $p_n(y_0,\ldots,y_n)$ of the observation sequence $Y_k = y_k$, $k \leq n$ is approximated using the unbiased particle estimate

$$\frac{1}{\sqrt{2\pi}^{(n+1)}} \prod_{0\leq k\leq n} \frac{1}{N} \sum_{1\leq i\leq N} \exp\left(-\frac{1}{2}(y_k - h_k(\xi^i_k))^2\right) \ \simeq_{N\uparrow\infty} \ p_n(y_0,\ldots,y_n)$$

In statistical machine learning and advanced signal processing the particle approximation (2.8) of these measures are called particle filters, or sequence Monte Carlo samplers. For a more thorough discussion on these models, we refer to section 10.2, section 10.1, and section 11.2.

We illustrate these particle models with the Bernoulli signal jump filtering problem discussed in the end of section 2.2.3. The resulting algorithm is also termed a particle filter.

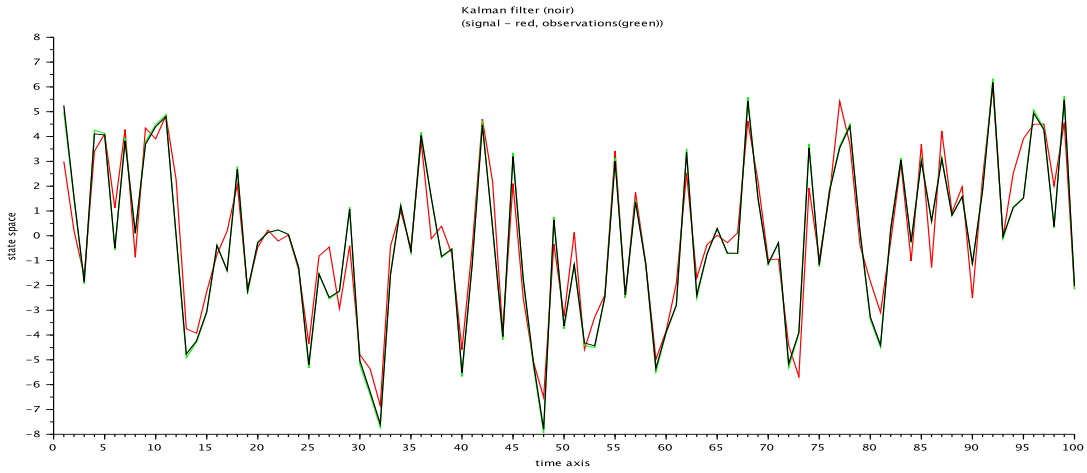
The next picture compares the extended Kalman filter with the smoothed empirical means of the genealogical tree associated with the particle filter.



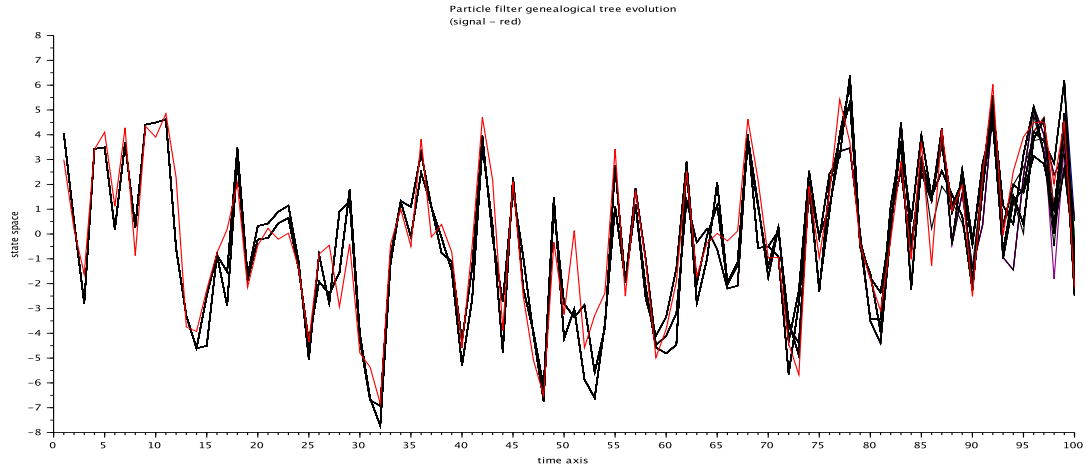
The next picture illustrate the extended Kalman filter associated with a Bernoulli switching signal

$$X_n = (\epsilon_n \times 1 + (1 - \epsilon_n) \times 1/2)\ X_{n-1} + W_n$$

with $\mathbb{P}(\epsilon_n = 1) = 1\%$, and a collection of uniform r.v. W_n on $[-5, 5]$, and an unit sensor perturbation noise.



The next figure illustrate the genealogical tree evolution associated with a particle filter with $N = 100$ particles.



We end this section with a discussion on the twisted models discussed in section 2.4.2. We consider a signal-observation model $(X_n, Y_n) \in \mathbb{R}^{p+q}$ of the form

$$\left\{ \begin{array}{rcl} X_n & = & A_n(X_{n-1}) + W_n \\ Y_n & = & C_n X_n + c_n + V_n \end{array} \right.$$

with independent centered Gaussian r.v. W_n and V_n with covariance matrices R_n^w and R_n^v. B_n, C_n and c_n are matrices and vectors with appropriate dimension, and A_n a collection of functions from $\mathbb{R}^p$ into itself. In this situation, using the standard Bayes' rule for Gaussian densities (cf. (8.29) we have

$$\text{Law}\,(X_n \mid X_{n-1},\ Y_n) = \mathcal{N}\left(\widehat{m}_n(X_{n-1}), \widehat{\Sigma}_n\right)$$

with the mean and covariance matrices given by

$$\begin{array}{rcl} \widehat{m}_n(X_{n-1}) & = & A_n(X_{n-1}) + R_n^w C_n' \Sigma_n(R_n^w)^{-1}\ (Y_n - (C_n A_n(X_{n-1}) + c_n)) \\ \widehat{\Sigma}_n & = & (I - R_n^w C_n' \Sigma_n(R_n^w)^{-1} C_n) R_n^w \quad \text{and} \quad \Sigma_n(R_n^w) := C_n R_n^w C_n' + R_n^v \end{array}$$

In addition, recalling that

$$\mathbb{P}\,(Y_n \in dy_n \mid X_{n-1}) = \int \mathbb{P}\,(Y_n \in dy_n \mid X_n = x_n)\ \mathbb{P}\,(X_n \in dx_n \mid X_{n-1})$$

we find that

$$\text{Law}\,(Y_n \mid X_{n-1}) = \mathcal{N}\left(C_n A_n(X_{n-1}), C_n R_n^w C_n' + R_n^v\right)$$

In this situation, the twisted Markov transition and the corresponding potential function given in (2.15) are defined by

$$\widehat{M}_n(x_{n-1}, dx_n) = \mathbb{P}\,(X_n \in dx_n \mid X_{n-1} = x_{n-1},\ Y_n = y_n)$$

and

$$\widehat{G}_n(x_n) \propto \exp\left(-\frac{1}{2}(y_{n+1} - C_{n+1}A_{n+1}(x_n))'[C_{n+1}R_{n+1}^w C_{n+1}' + R_{n+1}^v]^{-1}(y_{n+1} - C_{n+1}A_{n+1}(x_n))\right)$$

2.5.3 Sequential Monte Carlo methods

We consider a sequence of probability measures μ_n on some state space E defined in terms of the product of some functions h_n and some reference measure λ

$$\mu_n(dz) = \frac{1}{\mathcal{Z}_n}\left\{\prod_{1 \le p \le n} h_p(z)\right\}\ \lambda(dz)$$

We let $G_{n-1} = h_n$, and X_n a Markov chain with transitions M_n s.t. $\mu_n = \mu_n M_n$. In this situation, μ_n coincides with the n-th time marginal η_n of the Feynman-Kac measure $\mathbb{Q}_n$ given in (2.1). In addition, the normalizing constants are given by the product formulae

$$\mathcal{Z}_n/\mathcal{Z}_0 = \mathbb{E}\left(\prod_{0 \le k < n} G_k(X_k)\right) = \prod_{0 \le k < n} \eta_k(G_k)$$

We illustrate these models with a series of examples:

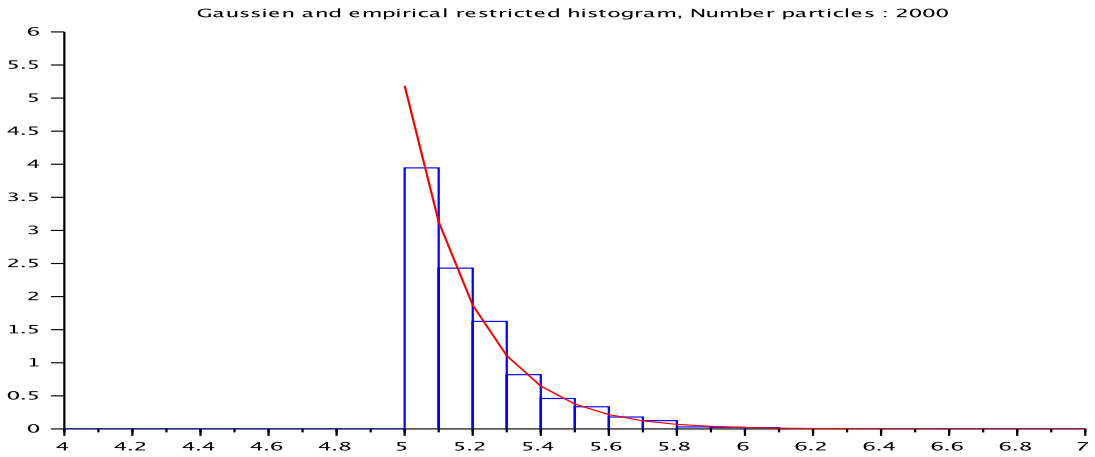
- *Bolztmann-Gibbs measures* associated with some inverse cooling schedule $\beta_n \uparrow$, with $\beta_0 = 0 = \beta_{-1}$ and some energy type function V:

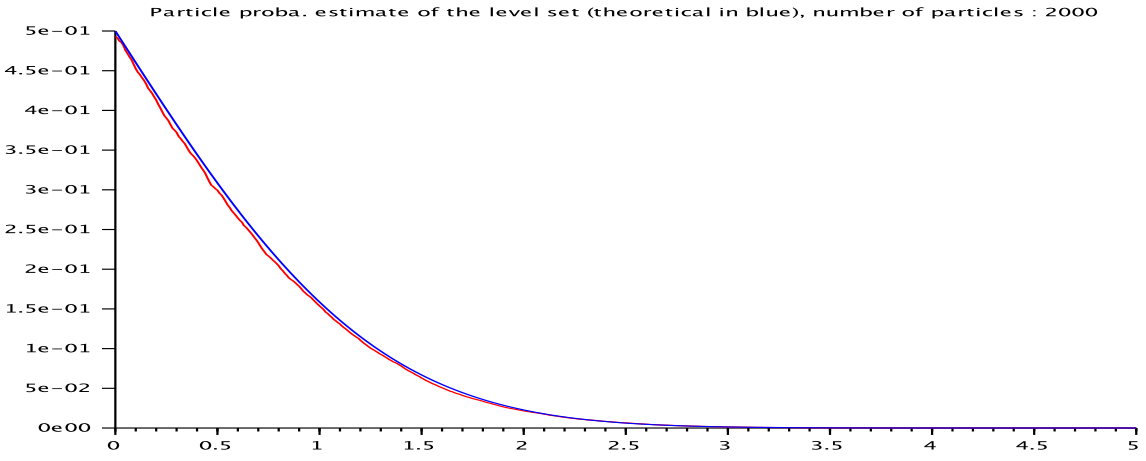
$$h_n = e^{-(\beta_n - \beta_{n-1})V} \implies \mu_n(dx) \propto e^{-\beta_n V(x)}\ \lambda(dx)$$

- *Bolztmann-Gibbs measures* associated with some decreasing subsets $A_n \uparrow$ and some probability measure λ:

$$h_n = 1_{A_n} \implies \mu_n(dx) \propto 1_{A_n}(x)\ \lambda(dx)$$

The following pictures illustrate the empirical histogram of a particle scheme with $N = 200$ particle, and the particle estimation of the normalizing constants in the case $\lambda = \mathcal{N}(0,1)$ and a terminal level set $A_n = [5, \infty[$.





- *Hidden Markov chain models* associated with a couple signal-observation processes (X_n, Y_n) depending on some r.v. Θ:

$$h_n(\theta) = p(y_n \mid \theta,\ y_0, \ldots, y_{n-1}) \implies \mu_n = \text{Law}\,(\Theta \mid y_0, \ldots, y_n)$$

In the above display we have used standard abusive but natural Bayesian notation. For instance, $p(y_n \mid \theta,\ y_0, \ldots, y_{n-1})$ stands for the density of the r.v. Y_n evaluated at $Y_n = y_n$ given $\Theta = \theta$ and $(Y_0, \ldots, Y_{n-1}) = (y_0, \ldots, y_{n-1})$.

- *Extended Hidden Markov chain models* are defined in terms of a realization $\xi = (\xi_n)_{n\geq 0}$ of the conditional particle approximation of the optimal prediction problem of X_n given $(Y_0, \ldots, Y_{n-1})$ and given the value of the parameter Θ. More precisely, we have that

$$h_n(\theta, \xi) = \int p(y_n \mid x_n, \theta, y_0, \ldots, y_{n-1}) \underbrace{\frac{1}{N} \sum_{1\leq i\leq N} \delta_{\xi_n^i}(dx_n)}_{\simeq dp(x_n|\theta, y_0, \ldots, y_{n-1})}$$

$$\Longrightarrow \text{ The } \Theta\text{-marginal of } \mu_n = \text{Law}\,(\Theta \mid y_0, \ldots, y_n)$$

Chapter 3

Some advanced Monte Carlo methodologies

3.1 Integro-partial differential equations

We let $a_t(x) = (a_{t,i}(x))_{1\leq i\leq p} \in \mathbb{R}^p$, $\lambda_t(x) \in \mathbb{R}_+$, and $\sigma_t(x) = (\sigma_{t,i,j}(x))_{1\leq i,j\leq p} \in \mathbb{R}^{p\times p}$ be a couple of smooth functions on $\mathbb{R}^p$. We also consider a smooth function $q_t(x,y)$ on $\mathbb{R}^{p+p}$ s.t. $\int q_t(x,y)dy = 1$, and we denote by $S_t(x,dy)$ the Markov transition

$$S_t(x,dy) = q_t(x,y)dy$$

We let $p(t,x)$ be the solution of the partial differential equation

$$\partial_t p_t = -\sum_{i=1}^d \partial_{x_i}\left(a_{t,i}\ p_t\right) + \frac{1}{2}\sum_{i,j=1}^d \partial_{x_i,x_j}\left(\left(\sigma_t(\sigma_t)^T\right)_{i,j}\ p_t\right) + \int p_t(y)\ \lambda_t(y)\ \left[q_t(y,x)\ dy - \delta_x(dy)\right] \quad (3.1)$$

This evolution model is sometimes called the Fokker-Planck equation. In the weak form, by a simple integration by part we have

$$\eta_t(f) = \int f(x)\ \eta_t(dx) \quad \text{with} \quad \eta_t(dx) = p_t(x)\ dx \implies \frac{d}{dt}\eta_t(f) = \eta_t\left(L_t(f)\right)$$

for sufficiently smooth test functions f with the infinitesimal generator

$$L_t(f)(x) := \underbrace{\sum_{i=1}^d a_{t,i}\ \partial_{x_i} f(x)}_{\text{drift term}} + \underbrace{\frac{1}{2}\sum_{i,j=1}^d \left(\sigma_t(\sigma_t)^T\right)_{i,j}\ \partial_{x_i,x_j} f(x)}_{\text{diffusion term}} + \underbrace{\lambda_t(x)\ \int [f(y)-f(x)]\ S_t(x,dy)}_{\text{jump term}}$$

We consider the jump-diffusion process X_t^λ with jump rate $\lambda_t(X_t)$ and evolving between two consecutive jump times according to the following stochastic differential equation

$$dX_t^\lambda = a_t\left(X_t^\lambda\right)\ dt + \sigma_t\left(X_t^\lambda\right)\ dW_t \quad (3.2)$$

where W_t stands for a p-dimensional standard Brownian motion. The jump times T_n are defined sequentially by setting $T_0 = 0$ and

$$\forall n \geq 1 \quad T_n = \inf\left\{t \geq T_{n-1} \text{ s.t. } \int_{T_{n-1}}^t \lambda_s(X_s^\lambda)ds \geq -\log U_n\right\}$$

where U_n stands for a sequence of uniform r.v. $U_n \sim \text{Unif}]0,1]$. Whenever p_0 is the density of the initial r.v. X_0^λ, we have

$$\forall t \geq 0 \qquad \mathbb{P}\left(X_t^\lambda \in dx\right) = p_t(x)\ dx$$

The analysis of these models is provided in section 5.1 and section 5.6. We choose a time mesh t_n with $(t_n - t_{n-1}) = \epsilon$ with intermediate time steps $(t_{n+1} - t_{n+\frac{1}{2}}) = \epsilon/2$, a sequence of i.i.d. Gaussian r.v. $V_n \sim \mathcal{N}(0, Id_{p\times p})$, a sequence of uniform r.v. $U_n \sim \text{Unif}[0,1]$, and a sequence of i.i.d. r.v. $F_n(x)$ with common distribution

$$\mathbb{P}\left(F_n(x) \in dy\right) = S_t(x, dy) \tag{3.3}$$

The ϵ-approximated Monte Carlo simulation of X_t is given by the recursion

$$X_{t_n}^{\lambda,\epsilon} \rightsquigarrow X_{t_{n+\frac{1}{2}}}^{\lambda,\epsilon} = X_{t_n}^{\lambda,\epsilon} + a_{t_n}(X_{t_n}^{\lambda,\epsilon})\ \epsilon + \sqrt{\epsilon}\ \sigma_{t_n}(X_{t_n}^{\lambda,\epsilon})\ V_n$$

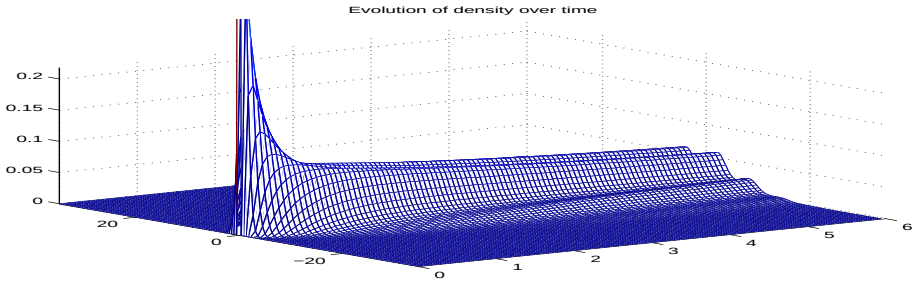
$$\rightsquigarrow X_{t_{n+1}}^{\lambda,\epsilon} = \begin{cases} X_{t_{n+\frac{1}{2}}}^{\lambda,\epsilon} & \text{if} \quad U_n \leq \exp\left\{-\lambda_{t_n}(X_{t_{n+\frac{1}{2}}}^{\lambda,\epsilon})\delta\right\} \\ F_n\left(X_{t_{n+\frac{1}{2}}}^{\lambda,\epsilon}\right) & \text{if} \quad U_n > \exp\left\{-\lambda_{t_n}(X_{t_{n+\frac{1}{2}}}^{\lambda,\epsilon})\delta\right\} \end{cases} \tag{3.4}$$

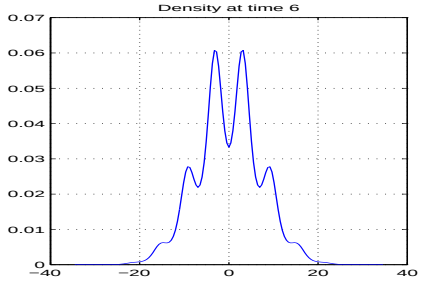
We also notice that $X_{t_{n+1}}^{0,\epsilon} = X_{t_{n+\frac{1}{2}}}^{0,\epsilon}$, for any $n \geq 0$ and any time step ϵ. In this situation, we also have

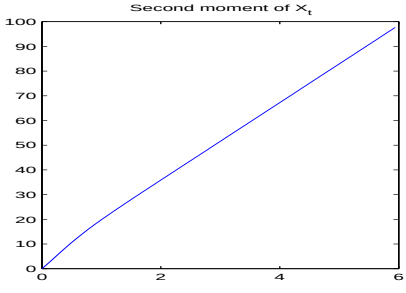
$$t_n = \lfloor t/\epsilon \rfloor\ \epsilon := n\epsilon \quad \Longrightarrow \mathbb{P}\left(X_{t_n}^{\lambda,\epsilon} \in dx\right) \simeq_{\epsilon \downarrow 0} p_t(x)\ dx$$

The next picture illustrates the solution $p_t(x)$ of the Fokker-Planck equation (3.1) associated with the pure diffusion process starting at the origin

$$dX(t) = \sin(X(t))dt + \sigma\ dB(t)$$







3.2 Mean field particle models

3.2.1 Discrete generation models

Suppose we are given a Markov chain X_n with transition probabilities K_n on some state spaces E_n. In this situation, the law η_n of the random states satisfies the evolution equation

$$\eta_n = \eta_{n-1}K_n \iff \eta_n(dx_n) = \int \eta_{n-1}(dx_{n-1})\; K_n(x_{n-1}, dx_n) \tag{3.5}$$

The sampling of this chain requires to sample random variables X_n with distribution $K_n(x_{n-1}, dx_n)$, for any given state $X_{n-1} = x_{n-1}$. If K_n is replaced by some Markov transition $K_{n,\eta_{n-1}}$ that depends on η_{n-1}, the sampling of the random states X_n requires to compute the law η_{n-1} of the random states X_{n-1}. Notice that in this situation the evolution of these distribution is given by a nonlinear system

$$\eta_n = \eta_{n-1}K_{n,\eta_{n-1}}$$

This shows that $\eta_n = \mathrm{Law}(\overline{X}_n)$, with the nonlinear Markov chain model

$$\mathbb{P}\left(\overline{X}_n \in dx_n \mid \overline{X}_{n-1} = x_{n-1}\right) = K_{n,\eta_{n-1}}(x_{n-1}, dx_n) \quad \text{with} \quad \eta_{n-1} = \mathrm{Law}(\overline{X}_{n-1})$$

In general, this type of nonlinear evolution equations cannot be solved explicitly, and we need to introduce another level of approximation. The mean field particle interpretation of these nonlinear Markov chain models is defined by an N-interacting particle Markov chain $\xi_n = \left(\xi_n^i\right)_{1\leq i\leq N}$ on the product space E_n^N with elementary transitions

$$\mathbb{P}\left(\xi_n \in \boldsymbol{dx_n} \mid \xi_{n-1}\right) = \prod_{1\leq i\leq N} K_{n,\eta_{n-1}^N}\left(\xi_{n-1}^i, dx_n^i\right) \quad \text{with} \quad \eta_{n-1}^N := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_{n-1}^i} \tag{3.6}$$

In the above display $\boldsymbol{dx_n} := d(x_n^1, \ldots, x_n^N)$ stands for an infinitesimal neighborhood of the state $\boldsymbol{x_n} = (x_n^i)_{1\leq i\leq N} \in E_n^N$. In other words, given ξ_{n-1}, the r.v. ξ_n^i are independent r.v. with distribution $K_{n,\eta_{n-1}}\left(\xi_{n-1}^i, dx_n\right)$. For a more thorough discussion on these discrete generation nonlinear processes, we refer the reader to section 9.1.

3.2.2 Continuous time models

The evolution equation (3.4) can be interpreted as the continuous time version of (3.5). To get some feasible solution, we have implicitly assumed in (3.3) that the values of the functions $a_t(x)$, $\lambda_t(x)$ and $\sigma_t(x)$ are known at any state $x \in \mathbb{R}^p$, and it is possible to sample r.v. with the distribution $q_t(x, y)dy$

In more general instance, these functions may depend on the solution p_t of the integro-partial differential equations (3.1). For instance, let us suppose that

$$a_t(x) = b_{t,\eta_t}(x) := \int \overline{a}_t(x, y)\; \eta_t(dy) \quad \text{and} \quad \sigma_t(x) = \tau_{t,\eta_t}(x) := \int \overline{\sigma}_t(x, y)\; \eta_t(dy) \tag{3.7}$$

and

$$\lambda_t(x) = V_{t,\eta_t}(x) := \int \overline{\lambda}_t(x, y)\; \eta_t(dy) \quad \text{and} \quad S_t(x, y) = \mathcal{S}_{t,\eta_t}(x, dy) = \left[\int \overline{q}_t((x, y), z)\; \eta_t(dz)\right] dy \tag{3.8}$$

with $\eta_t(dx) = p_t(x)\; dx$, and for some regular function $\overline{a}_t$, $\overline{\sigma}_t$, $\overline{\lambda}_t$ and $\overline{q}_t$. In this situation, (3.1) is a nonlinear partial differential equation. In addition, we have

$$\frac{d}{dt}\eta_t(f) = \eta_t\left(L_{t,\eta_t}(f)\right)$$

with the infinitesimal generator

$$L_{t,\eta_t}(f)(x) := \underbrace{\sum_{i=1}^{d} b_{t,\eta_t,i}(x)\ \partial_{x_i} f(x)}_{\text{drift term}} + \underbrace{\frac{1}{2}\sum_{i,j=1}^{d} \left(\tau_{t,\eta_t}(\tau_{t,\eta_t})^T\right)_{i,j}(x)\ \partial_{x_i,x_j} f(x)}_{\text{diffusion term}} + \underbrace{V_{t,\eta_t}(x)\ \int [f(y)-f(x)]\ \mathcal{S}_{t,\eta_t}(x,dy)}_{\text{jump term}}$$

The mean field particle interpretation of this model is defined by a Markov chain $\xi_t = \left(\xi_t^i\right)_{1\leq i\leq N}$ on the product space $(\mathbb{R}^d)^N$ with infinitesimal generator defined, for sufficiently regular functions F on $(\mathbb{R}^d)^N$, by the following formulae

$$\mathcal{L}_t(F)(x^1,\ldots,x^N) := \sum_{1\leq i\leq N} L^{(i)}_{t,m(x)}(F)(x^1,\ldots,x^i,\ldots,x^N) \quad \text{with} \quad m(x) := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{x^i} \tag{3.9}$$

In the above display, $L^{(i)}_{t,m(x)}$ stands for the operator $L_{t,m(x)}$ acting on the function $x^i \mapsto F(x^1,\ldots,x^i,\ldots,x^N)$. For a more thorough discussion on these continuous time particle models we refer the reader to section 9.3.2.

The stochastic model (3.4) and its mean field interpretation (3.9) cannot be simulated on some computer without an additional level of approximation.

The discrete time mean field particle approximated scheme is defined by an N-interacting jump particle process on the product space $(\mathbb{R}^p)^N$ with a two step elementary transitions

$$\xi^{\lambda,\epsilon}_{t_n} := \left(\xi^{\lambda,\epsilon,i}_{t_n}\right)_{1\leq i\leq N} \longrightarrow \xi^{\lambda,\epsilon}_{t_{n+1/2}} := \left(\xi^{\lambda,\epsilon,i}_{t_{n+1/2}}\right)_{1\leq i\leq N} \longrightarrow \xi^{\lambda,\epsilon}_{t_{n+1}} = \left(\xi^{\lambda,\epsilon,i}_{t_{n+1}}\right)_{1\leq i\leq N}$$

To clarify the presentation, when there is no confusion we drop the indices $(.)^{(\lambda,\epsilon)}$, and we write $\xi^i_{t_n}$ instead of $\xi^{\lambda,\epsilon,i}_{t_n}$.

In this notation, replacing in the evolution equation (3.4) the measure $p_{t_n}(y)dy$ by the empirical measures

$$m(\xi_{t_n}) := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi^i_{t_n}}$$

the transition $\xi^i_{t_n} \rightsquigarrow \xi^i_{t_{n+1/2}}$ of the i-th particle takes the form

$$\xi^i_{t_{n+1/2}} = \xi^i_{t_n} + \underbrace{\frac{1}{N}\sum_{1\leq j\leq N} \overline{a}_{t_n}(\xi^i_{t_n},\xi^j_{t_n})}_{=b_{t_n,m(\xi_{t_n})}(\xi^i_{t_n})}\ \epsilon + \sqrt{\epsilon}\ \underbrace{\frac{1}{N}\sum_{1\leq j\leq N} \overline{\sigma}_{t_n}(\xi^i_{t_n},\xi^j_{t_n})}_{=\tau_{t_n,m(\xi_{t_n})}(\xi^i_{t_n})}\ V^i_n$$

where $(V^i_n)_{1\leq i\leq N}$ stands for N i.d.d. copies of V_n.

In much the same way, the the transition $\xi^i_{t_{n+1/2}} \rightsquigarrow \xi^i_{t_n}$ of the i-th particle takes the form

$$\xi^i_{t_{n+1}} = \begin{cases} \xi^i_{t_{n+1/2}} & \text{if} \quad U^i_n \leq \exp\left\{-\lambda^N_{t_n}\left(\xi^i_{t_{n+1/2}}\right)\delta\right\} \\ F^N_n\left(\xi^i_{t_{n+1/2}}\right) & \text{if} \quad U^i_n > \exp\left\{-\lambda^N_{t_n}\left(\xi^i_{t_{n+1/2}}\right)\delta\right\} \end{cases}$$

with N i.d.d. copies $(U_n^i)_{1\leq i\leq N}$ of U_n, $F_n^N\left(\xi^i_{t_{n+\frac{1}{2}}}\right)$ a r.v. with distribution

$$\mathcal{S}_{t_n,m\left(\xi_{t_{n+1/2}}\right)}\left(\xi^i_{t_{n+1/2}},dy\right)=\frac{1}{N}\sum_{1\leq j\leq N}\overline{q}_{t_n}\left(\left(\xi^i_{t_{n+1/2}},y\right),\xi^j_{t_{n+1/2}}\right)\,dy$$

and

$$\lambda^N_{t_n}\left(\xi^i_{t_{n+1/2}}\right):=V_{t_n,m\left(\xi_{t_{n+1/2}}\right)}\left(\xi^i_{t_{n+1/2}}\right)=\frac{1}{N}\sum_{1\leq j\leq N}\overline{\lambda}_{t_n}\left(\xi^i_{t_{n+1/2}},\xi^j_{t_{n+1/2}}\right)$$

In this case, under some rather weak regularity conditions we prove that

$$t_n=\lfloor t/\epsilon\rfloor\,\epsilon:=n\epsilon\quad\Longrightarrow\frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi^i_{t_n}}\simeq_{N\uparrow\infty}\mathbb{P}\left(X^{\lambda,\epsilon}_{t_n}\in dx\right)\simeq_{\epsilon\downarrow 0}\;p_t(x)\;dx$$

3.3 Some illustrations

3.3.1 Discrete generation Feynman-Kac models

We consider the Feynman-Kac measures η_n discussed in (2.2). To simplify the presentation we assume that the potential functions G_n take values in $[0,1]$. In this situation, combining (2.3) with (8) and (9) we readily check that

$$\eta_n=\Psi_{G_{n-1}}(\eta_{n-1})M_n=\eta_{n-1}K_{n,\eta_{n-1}}$$

with

$$K_{n,\eta_{n-1}}(x_{n-1},dx_n):=G_{n-1}(x_{n-1})\;M_n(x_{n-1},dx_n)+(1-G_{n-1}(x_{n-1}))\;\left(\Psi_{G_{n-1}}(\eta_{n-1})M_n\right)(dx_n)$$

The N-particle interpretation (3.6) of the Feynman-Kac measures η_n is defined by an N-interacting particle Markov chain $\xi_n=\left(\xi^i_n\right)_{1\leq i\leq N}$ on the product space E_n^N. Given the values of the N particle model ξ_{n-1} at time $(n-1)$, the N particles ξ^i_n are independent r.v. with distributions

$$K_{n,\eta^N_{n-1}}\left(\xi^i_{n-1},dx^i_n\right)$$

$$=G_{n-1}(\xi^i_{n-1})\;M_n(\xi^i_{n-1},dx^i_n)+\left(1-G_{n-1}(\xi^i_{n-1})\right)\;\sum_{1\leq j\leq N}\frac{G_{n-1}(\xi^j_{n-1})}{\sum_{1\leq k\leq N}G_{n-1}(\xi^k_{n-1})}\;M_n(\xi^j_{n-1},dx^i_n)$$

with $1\leq i\leq N$. The resulting particle algorithm coincides with the fixed population size branching process discussed in section 2.3.2. Further details on particle Feynman-Kac models are provided in section 9.1.4 and in chapter 10.

3.3.2 A particle systemic risk model

We consider the log-monetary reserves $(X^i_t)_{1\leq i\leq N}$ of N banks. The inter-bank exchanges (borrowing and lending) are represented by the diffusion equation

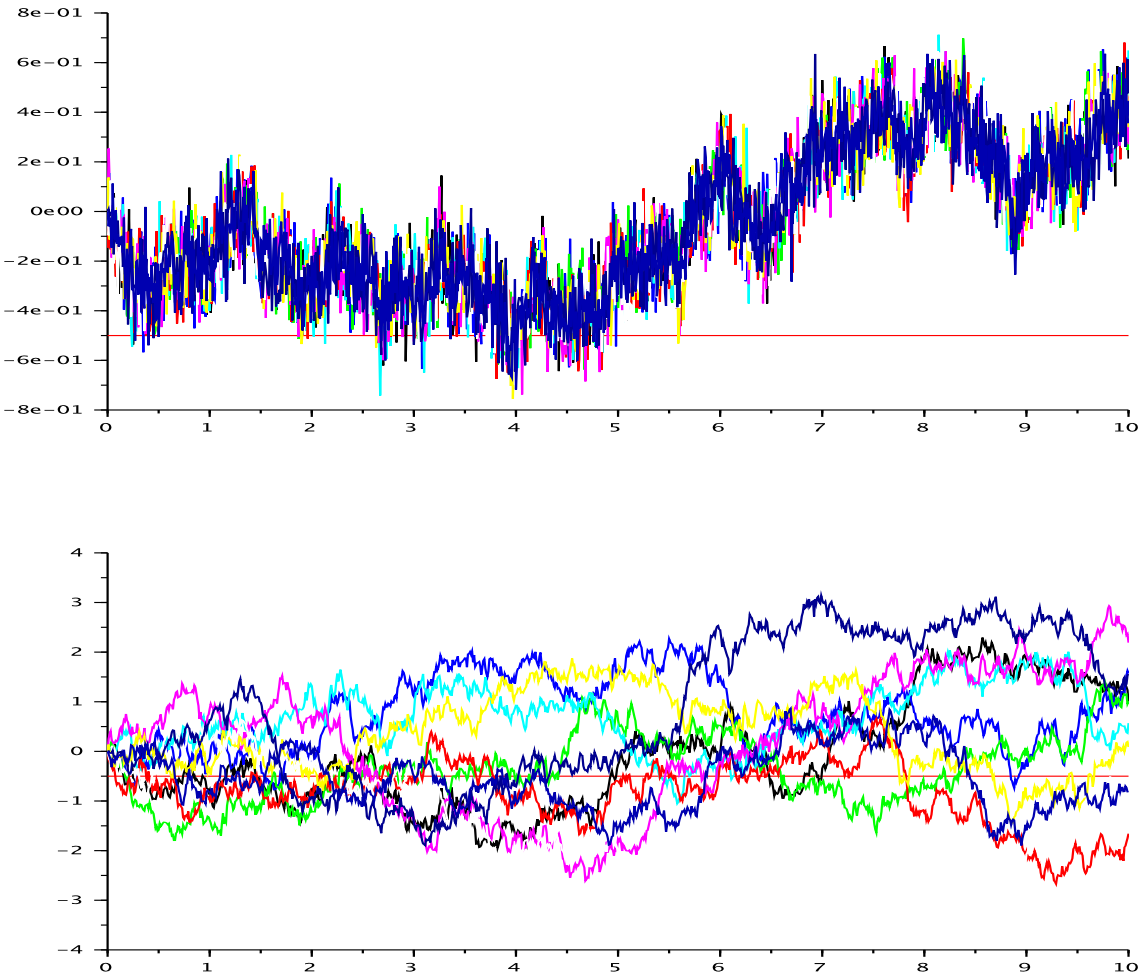
$$dX^i_t=\frac{\alpha}{N}\sum_{1\leq j\leq N}(X^j_t-X^i_t)\;dt+\sigma\;dW^i_t$$

where $(W^i_t)_{1\leq i\leq N}$ stands for N independent Brownian motions, and a couple of parameters α and σ. This model is the mean field approximation of the nonlinear process associated with the parameters (3.7) and (3.8) with

$$\overline{a}_t(x,y)=\alpha\times(y-x)\qquad\sigma_t(x)=\sigma\quad\text{and}\quad\lambda_t(x)=0$$

This model has been introduced by J.P. Fouque and L.H. Sun in [273] (see also [105] for a mean field game interpretation of this model). Simulations shows that stability is created by increasing the parameter α. Nevertheless the systemic risk is also increase when α is large.

The following pictures illustrates an Euler type discrete time approximation of the model with 10 banks when $\alpha = 100$ and $\alpha = 0$.



3.3.3 The burgers equation

We consider the nonlinear model (3.7) and (3.8) on $\mathbb{R}$ with

$$\overline{a}_t(x,y) = 1_{[x,\infty[}(y) \qquad \sigma_t(x) = \sigma > 0 \quad \text{and} \quad \lambda_t(x) = 0$$

We set

$$V_t(x) = \int_x^\infty p_t(y)dy = a_t(x) \implies \partial_x V_t = -p_t$$

This implies that

$$\begin{aligned}\partial_t p_t &= -\partial_t\partial_x V_t = -\partial_x\ \partial_t V_t \\ &= -\partial_x\left(V_t\ p_t\right) + \frac{\sigma^2}{2}\partial_{x,x}p_t = \partial_x\left[V_t\ \partial_x V_t\ -\frac{\sigma^2}{2}\ \partial_{x,x}V_t\right] \qquad \text{(by (3.1))}\end{aligned}$$

In other words, V_t satisfies the Burgers equation

$$\partial_t V_t = -V_t\ \partial_x V_t\ + \frac{\sigma^2}{2}\ \partial_{x,x}V_t$$

An explicit solution (cf. exercise 5.7.1 in [211]) is given by

$$V_t(x) = \frac{\mathbb{E}\left(1_{]-\infty,0]}(x+\sigma W_t)\ e^{-\frac{1}{\sigma^2}(x+\sigma W_t)}\right)}{\mathbb{E}\left(1_{]0,\infty[}(x+\sigma W_t)\right) + \mathbb{E}\left(1_{]-\infty,0]}(x+\sigma W_t)\ e^{-\frac{1}{\sigma^2}(x+\sigma W_t)}\right)} \tag{3.10}$$

This formula can be computed using a crude Monte Carlo technique or using the function $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\ \int_x^\infty e^{-y^2}dy$, with

$$V_t(x) = \frac{\exp\left(\frac{t-2x}{2\sigma^2}\right)\ \left(1-\frac{1}{2}\mathrm{erfc}\left(\frac{t-x}{\sqrt{2\sigma^2 t}}\right)\right)}{\exp\left(\frac{t-2x}{2\sigma^2}\right)\ \left(1-\frac{1}{2}\mathrm{erfc}\left(\frac{t-x}{\sqrt{2\sigma^2 t}}\right)\right) + \frac{1}{2}\mathrm{erfc}\left(-\frac{x}{\sqrt{2\sigma^2 t}}\right)} \tag{3.11}$$

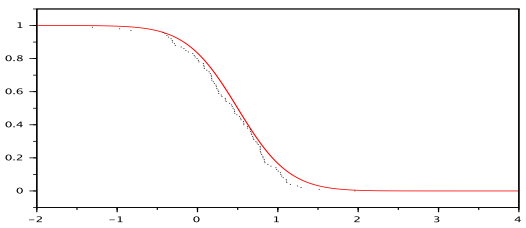
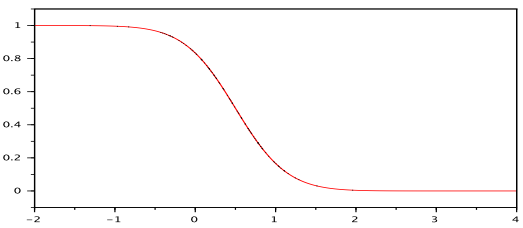
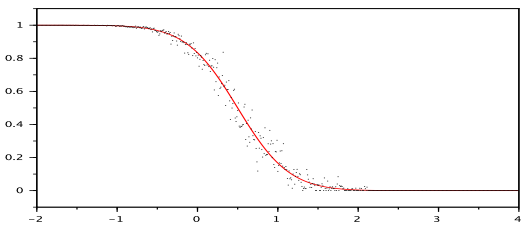
The mean field particle interpretation of this model is given by

$$dX_t^i = \frac{1}{N}\sum_{1\leq j\leq N} 1_{[X_t^i,\infty[}(X_t^j)\ dt + \sigma\ dW_t^i$$

where $(W_t^i)_{1\leq i\leq N}$ stands for N independent Brownian motions. In this case, we have

$$V_t^N(x) := \frac{1}{N}\sum_{1\leq j\leq N} 1_{[x,\infty[}(X_t^j)\ \simeq_{N\uparrow\infty}\ V(x)$$

The following pictures illustrates these three approximations. The top l.h.s picture compares the exact solution with the mean field particle estimate based on an Euler type scheme with $N = 100$ particles and a $\Delta t = .01$ time step. The top r.h.s. represents the exact values on the simulated states, and the bottom picture compares the crude Monte Carlo method with the exact solution.

3.3.4 Langevin-McKean-Vlasov processes

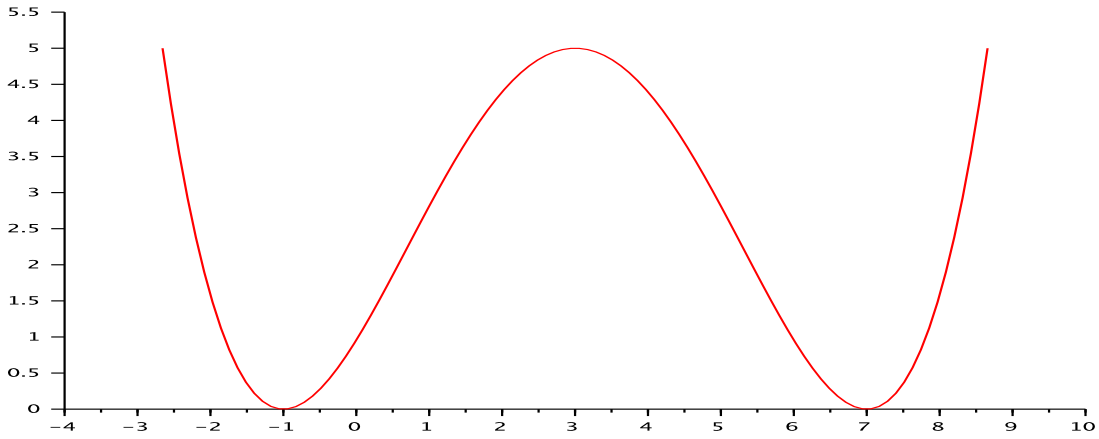
The Langevin-McKean-Vlasov model is a stochastic gradient process on $\mathbb{R}$ associated with some smooth energy function V, coupled with an attraction or repulsion force around ensemble averages. This model is defined by N interacting diffusion processes

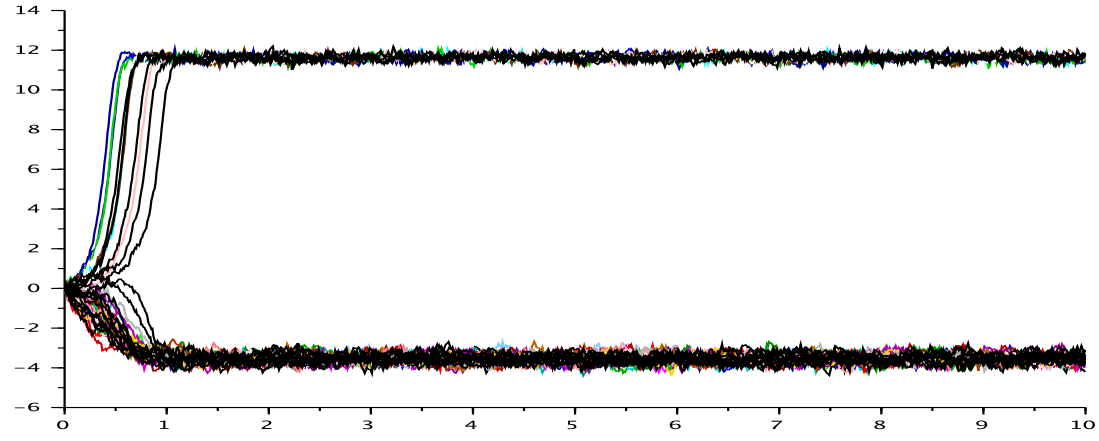
$$dX^i(t) = -\beta\ V'(X_t^i)dt + \alpha\ \left(\frac{1}{N}\sum_{1\leq j\leq N} X_t^j - X_t^i\right)\ dt + \sigma\ dW_t^i$$

where $(W_t^i)_{1\leq i\leq N}$ stands for N independent Brownian motions, and some fixed parameters α, β and σ. This model has been introduced by S. Herrmann, and J. Tugaut in [342]. This stochastic model coincides with the mean field particle interpretation of the nonlinear model (3.7) and (3.8) on $\mathbb{R}$ with

$$\overline{a}_t(x,y)\ = -\beta\ V'(x) + \alpha \times (y-x) \qquad \sigma_t(x) = \sigma > 0 \quad \text{and} \quad \lambda_t(x) = 0$$

The following pictures illustrate the time evolution Langevin-McKean-Vlasov model with a double well potential with $\alpha = -7$, and $N = 50$ particles.





3.3.5 The Dyson equation

In nuclear physics, the statistical properties of the spectrum of quantum systems can be analyzed using the nonlinear Dyson equations

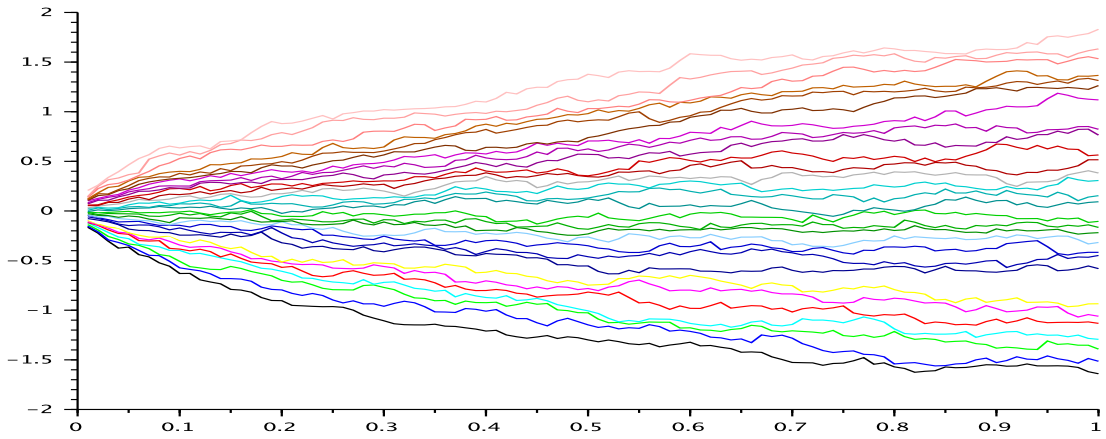
$$\begin{cases} d\lambda_i(t) = \dfrac{1}{N} \displaystyle\sum_{j \neq i} \frac{1}{\lambda_i(t) - \lambda_j(t)}\, dt + \sqrt{\frac{2}{N}}\, dW_t^i \\ 1 \leq i \leq N \end{cases}$$

with some initial conditions $\lambda_1 < \ldots < \lambda_N$. This mean field type particle model is slightly different from the nonlinear processes discussed in these lectures. One can show (cf. exercises 5.3.2-5.3.4 in [211]) that $\lambda_1 < \ldots < \lambda_N$ coïncides with the eigenvalues of the symmetric Gaussian matrices

$$A_{i,i}(t) = W^i(t)/\sqrt{N/2} \quad \text{et} \quad A_{i,j}(t) = A_{j,i}(t) = W^{i,j}(t)/\sqrt{N}$$

In the above display $W_{i,j}$, $1 \leq i < j \leq N$ and W_t^i, $1 \leq i \leq N$, stand for $N(N+1)/2$ independent Brownian motions on the real line.

The following picture illustrates the time evolution of $N = 30$ eigenvalues of the matrices $A(t)$ on the interval $[0, 1]$.



3.4 Boltzmann-Gibbs measures and Langevin diffusions

3.4.1 Stochastic gradient models

A stochastic process of particular interest is the pure diffusion model associated with $\lambda_t = 0$, $\sigma_{t,i,j}(x) = 1_{i=j}\sqrt{2}$, and the gradient $a_t = -\beta\, \nabla V$ of some smooth function V on $\mathbb{R}^p$, and some inverse temperature parameter $\beta \in \mathbb{R}_t$. We let μ_β be the Boltzmann-Gibbs measure on $\mathbb{R}^p$ defined by

$$\mu_\beta(dx) = \frac{1}{\mathcal{Z}_\beta}\, e^{-\beta V(x)}\, dx \quad \text{with} \quad \mathcal{Z}_\beta = \int e^{-\beta V(x)}\, dx$$

Several examples of Boltzmann-Gibbs measures are provided in section 7.1. In this situation, we have

$$\text{Law}(X_0) = \mu_\beta \implies \forall t \geq 0 \quad \text{Law}(X_t) = \mu_\beta \tag{3.12}$$

and for any bounded function f we have

$$\frac{1}{t}\int_0^t f(X_s)\ ds = \int f(x)\ \left[\frac{1}{t}\int_0^t \delta_{X_s}(dx)\right]\ ds \simeq_{t\uparrow\infty} \int f(x)\ \mu_\beta(dx)$$

These gradient flow models are discussed in section 7.8.

3.4.2 Metropolis-Hasting adjustments

Unfortunately, when $\lambda_t = 0$ and $a_t = -\beta\ \nabla V$, the discrete time version of (3.12) is not met. Nevertheless, we can recover this important property adding a Metropolis-Hasting type acceptance rate. The resulting algorithm is called the Metropolis-Hasting-Adjusted-Langevin model, and it is discussed in section 7.9. This model is defined by replacing in (3.4) the definition of $X^{0,\epsilon}_{t_{n+1}}$ by

$$X^{0,\epsilon}_{t_{n+1}} := \begin{cases} X^{0,\epsilon}_{t_{n+\frac{1}{2}}} & \text{if} \quad U_n \leq a_\epsilon\left(X^{0,\epsilon}_{t_n}, X^{0,\epsilon}_{t_{n+\frac{1}{2}}}\right) \\ \\ X^{0,\epsilon}_{t_n} & \text{if} \quad U_n > a_\epsilon\left(X^{0,\epsilon}_{t_n}, X^{0,\epsilon}_{t_{n+\frac{1}{2}}}\right) \end{cases}$$

with the acceptance rare

$$a_\epsilon\left(x,y\right) = 1 \wedge \left(e^{-\beta(V(y)-V(x))} \times \frac{p_\epsilon(y,x)}{p_\epsilon(x,y)}\right)$$

and the density function

$$p_\epsilon(x,y) = \frac{1}{(4\pi\epsilon)^{p/2}}\ \exp\left(-\frac{1}{4\epsilon}\left\|y - x + \beta\ \nabla V(x)\epsilon\right\|^2\right)$$

In this situation, we recover the fact that

$$\text{Law}(X_0) = \mu_\beta \implies \forall n \geq 0 \quad \text{Law}(X^\epsilon_{t_n}) = \mu_\beta$$

and for any bounded function f we have

$$\frac{1}{n}\sum_{0\leq k<n} f(X_k)\ \simeq_{n\uparrow\infty} \int f(x)\ \mu_\beta(dx)$$

3.4.3 Boltzmann-Gibbs measures on manifolds

We let $S = \varphi^{-1}(0)$ be some non empty and connected manifold defined in terms of the null level sets of some smooth function $\varphi = (\varphi_i)_{1\leq i\leq q}$ from $\mathbb{R}^{p+q}$ into $\mathbb{R}^q$. For instance, the 2-Torus is the null level set of the function

$$\varphi(x) = \left(R - \sqrt{x_1^2 + x_2^2}\right)^2 + x_3^2 - r^2$$

with $r < R$.

We assume that, for any $x \in S$, the vector space generated by the gradient vectors $\partial\varphi_i(x)$ has dimension q. We consider the $(q \times q)$-matrices

$$g_\perp^{-1} = \left(g_\perp^{i,j}\right)_{1\leq i,j\leq q} \quad \text{with} \quad g_\perp = (g_{\perp,i,j})_{1\leq i,j\leq q} = (\langle\partial\varphi_i, \partial\varphi_j\rangle)_{1\leq i,j\leq q}$$

The projection $\pi(x)(W(x))$ of a vector field $x \mapsto W(x) \in \mathbb{R}^{r=p+q}$ on the tangent space $T_x(S)$ of the manifold S at x is defined by the formula

$$\pi(x)(W(x)) = W(x) - \sum_{1\leq i\leq q}\left\langle\sum_{1\leq j\leq q} g_\perp^{i,j}(x)\partial\varphi_j(x), W(x)\right\rangle\ \partial\varphi_i(x)$$

We refer the reader to section 6.1 for a review of projection operators on vector spaces. In the above displayed formulae, $\langle ., . \rangle$ stands for the standard Euclidian inner product on $\mathbb{R}^q$.

We consider the diffusion equation

$$dX_t = -\pi(X_t)(\partial V)(X_t)\ dt + \left[\pi(X_t)\ dB_t - \frac{1}{2}\ \mathbb{H}(X_t)\ dt\right] \tag{3.13}$$

with a standard r-dimensional Brownian motion B_t, and the mean curvature vector field $\mathbb{H}$ defined by

$$\mathbb{H} = \sum_{1 \leq i \leq q} \mathrm{div}_\perp \left(\sum_{1 \leq j \leq q} g_\perp^{i,j}\ \partial \varphi_j \right)\ \partial \varphi_i$$

with

$$\mathrm{div}_\perp (W) = \frac{1}{\sqrt{\det(g_\perp)}} \sum_{1 \leq m \leq r} \partial_{x_m} \left(\sqrt{\det(g_\perp)}\ W^m \right)$$

In this situation, we have

$$X_0 \in S \implies \forall t \geq 0 \quad X_t \in S$$

We let η be the Boltzmann-Gibbs measure on S defined by

$$\eta(dx) := \frac{1}{\mathcal{Z}}\ e^{-2V(x)}\ \mu_S(dx)$$

where μ_S stands for the volume measure on S (cf. section 6.6.3 for a detailed discussion on these measures and an overview of integration techniques on manifolds). By construction, we have

$$\mathbb{P}\left(X_0 \in dx\right) := \implies \forall t \geq 0 \quad \mathbb{P}\left(X_t \in dx\right) := \frac{1}{\mathcal{Z}}\ e^{-2V(x)}\ \mu_S(dx)$$

In addition, for any bounded function f on S we have

$$\frac{1}{t} \int_0^t f(X_s)\ ds = \int f(x)\ \left[\frac{1}{t} \int_0^t \delta_{X_s}(dx)\right]\ ds \simeq_{t \uparrow \infty} \int f(x)\ \eta(dx)$$

When $V = 0$, the stochastic differential equation (3.13) reduces to the Brownian motion on the Torus. For a more thorough discussion on diffusions on manifolds, we refer the reader to section 6.4 and section 7.8.3.

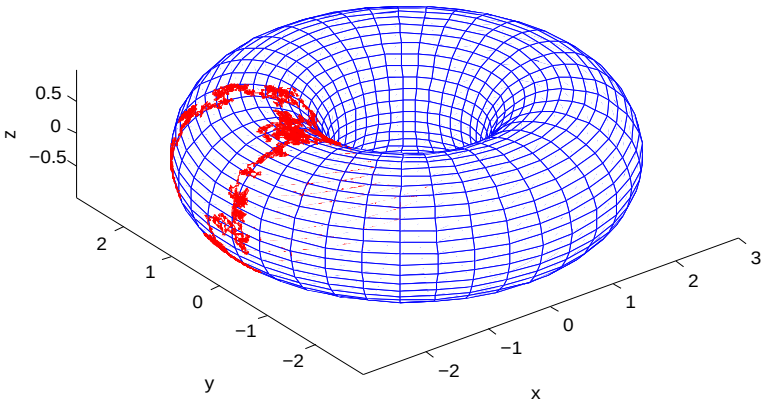
In practice, the sampling of the diffusion process (3.13) requires some discrete time approximation. For instance, an Euler type approximation on a time mesh $(t_n)_{n \geq 0}$ with $(t_n - t_{n-1}) = \epsilon \simeq 0$ is given by the equation

$$\begin{aligned} X_{t_n}^\epsilon - X_{t_{n-1}}^\epsilon \quad = \quad & -\pi(X_{t_n})(\partial V)(X_{t_n})\ (t_n - t_{n-1}) \\ & -\frac{1}{2}\ \mathbb{H}(X_{t_{n-1}}^\epsilon)\ (t_n - t_{n-1}) + \pi(X_{t_{n-1}}^\epsilon)\ \sqrt{t_n - t_{n-1}}\ \overline{B}_n \end{aligned}$$

where B_n stands for a sequence of i.i.d. centered and normalized Gaussian r.v. on $\mathbb{R}^r$. Unfortunately any type of these scheme ensure that $X_{t_n}^\epsilon$ stay in the Manifold S. As for deterministic dynamical systems, we often handle this issue by projecting each step on the manifold

$$\begin{aligned} X_{t_n}^\epsilon \quad = \quad & \mathrm{proj}_S \Big(X_{t_{n-1}}^\epsilon - \pi(X_{t_n})(\partial V)(X_{t_n})\ (t_n - t_{n-1}) \\ & -\frac{1}{2}\ \mathbb{H}(X_{t_{n-1}}^\epsilon)\ (t_n - t_{n-1}) + \pi(X_{t_{n-1}}^\epsilon)\ \sqrt{t_n - t_{n-1}}\ \overline{B}_n \Big) \end{aligned} \tag{3.14}$$

Another strategy is to use a description of the stochastic process in some judicious chart space. An illustration of a realization of a Brownian motion on the Torus is provided in the next picture.



3.4.4 Langevin equation in Riemannian manifolds

Suppose we are given a smooth parametrization of the manifold S

$$\psi \ : \ \theta \in S_\psi \subset \mathbb{R}^p \mapsto \psi(\theta) = \left(\psi^1(\theta), \ldots, \psi^r(\theta)\right)^T \in S \subset \mathbb{R}^r \tag{3.15}$$

with a well defined smooth inverse mapping $\phi = \psi^{-1}$. In differential geometry ϕ is called a chart or a coordinate mapping. These chart maps and their inverse $\psi = \phi^{-1}$ are often defined locally on open neighborhoods of each states and the set of all chart maps is often called an atlas (cf. section 6.3). For instance, the 2-Torus can be parametrized by the spherical coordinates

$$\psi\left(\theta\right) = \begin{pmatrix} \left(R + r\cos(\theta_1)\right)\cos(\theta_2) \\ \left(R + r\cos(\theta_1)\right)\sin(\theta_2) \\ r\sin(\theta_1) \end{pmatrix}$$

We let $g = (g_{i,j})_{1 \leq i,j \leq p}$ be the $(p \times p)$-matrix field on S_ψ defined by

$$\forall 1 \leq i, j \leq p \qquad g_{i,j} := \left\langle \partial_{\theta_i}\psi, \partial_{\theta_j}\psi \right\rangle \tag{3.16}$$

We let η be the Boltzmann-Gibbs measure on S_ψ defined by

$$\eta(d\theta) = \frac{1}{\mathcal{Z}}\ e^{-2U(\theta)}\ \sqrt{\det(g(\theta))}\ d\theta \quad \text{with} \quad U := V \circ \psi$$

We let $\overline{B}_t$ be the d-dimensional Brownian motion on the Riemannian manifold defined for any $1 \leq i \leq p$ by

$$d\overline{B}^i_t = \sum_{1 \leq k \leq p} \sqrt{g^{-1}}^{\,i}_{\,k}(\Theta_t)\ dB^k_t + \frac{1}{2}\sum_{1 \leq j \leq p} \frac{1}{\sqrt{\det(g(\Theta_t))}}\ \partial_{\theta_j}\left(\sqrt{\det(g)}\ g^{i,j}\right)(\Theta_t)\ dt$$

where $\sqrt{g^{-1}}^{\,i}_{\,k}$ stands for the (i,k)-th entry of the square root matrix of g^{-1}, and B_t stands for a standard p-dimensional Brownian motion. We consider the diffusion equation

$$d\Theta_t = -\nabla_g U(\Theta_t)\ dt + d\overline{B}_t \tag{3.17}$$

with the Riemannian gradient

$$\nabla_g U := \begin{bmatrix} \sum_{1 \leq j \leq p} g^{1,j}\ \partial_{\theta_j} U \\ \vdots \\ \sum_{1 \leq j \leq p} g^{p,j}\ \partial_{\theta_j} U \end{bmatrix} \quad \text{and} \quad g^{-1} = (g^{i,j})_{1 \leq i,j \leq p}$$

We refer the reader to section 6.3.2 and section 7.8.4 for a detailed discussion on Riemannian gradients and Riemannian Langevin diffusions. By construction, we have

$$X_t = \psi(\Theta_t) \quad \text{satisfies the Langevin diffusion (3.13)}$$

In addition, for any bounded function f on S_ψ we have

$$\frac{1}{t}\int_0^t f(\Theta_s)\, ds = \int f(\theta)\, \left[\frac{1}{t}\int_0^t \delta_{\Theta_s}(d\theta)\right] ds \simeq_{t\uparrow\infty} \int f(\theta)\, \eta(d\theta)$$

3.5 Feynman-Kac and Schrödinger equations

3.5.1 A particle absorption interpretation

We replace the r.v. $F_n(x)$ defined in (3.3) by an auxiliary cemetery state c; that is, we set $F_n(x) = c$. The resulting jump-diffusion process (3.2), resp. (3.4) can be interpreted as a particle model absorbed at rate $\lambda_t := V_t$, resp. at rate $\lambda_{t_n} := V_{t_n}$.

We let T, resp. T^ϵ, the first time the process (3.2), resp. (3.4) is placed in the cemetery state c. In this notation, for any bounded function f, we have

$$\begin{aligned}\mathbb{E}\left(f(X_t)\ 1_{T>t}\right) &= \mathbb{E}\left(f(X_t^0)\ \exp\left\{-\int_0^t V_s(X_s^0)ds\right\}\right)\\ &\simeq_{\epsilon\downarrow 0} \mathbb{E}\left(f(X_{t_n}^{0,\epsilon})\ \prod_{0\leq k<n} G_k\left(X_{t_k}^{0,\epsilon}\right)\right) = \mathbb{E}\left(f(X_{t_n}^{\lambda,\epsilon})\ 1_{T^\epsilon>t_n}\right)\end{aligned} \tag{3.18}$$

with the potential functions

$$G_k(x) = \exp\left\{-V_{t_k}(x)\epsilon\right\}$$

The above Feynman-Kac formulae are studied in section 4.2 and section 5.5. The particle interpretation of these models are defined as in (2.8). In computational physics, these particle algorithms are also termed Resampled and/or Quantum Monte Carlo methods (cf. for instance [93, 94]).

Under some regularity conditions, when $a_t = 0$ and $\sigma_t = Id$, the identity matrix, we have

$$\mathbb{P}\left(X_t \in dx\ ;\ T > t\right) = q_t(x)\ dx \tag{3.19}$$

for some density q_t satisfying the Schrödinger (imaginary time) equation

$$\partial_t q_t = L^V(q_t) := \frac{1}{2}\sum_{1\leq i\leq p} \partial_{x_i}^2 q_t - V_t(x) q_t(x)$$

This equation is sometimes written in terms of the Hamiltonian operator

$$\begin{aligned}&\mathcal{H} := -L^V = -L + V = -\tfrac{1}{2}\ \textstyle\sum_{1\leq i\leq p}\partial_{x_i}^2 + V_t\\ &\Longleftrightarrow \forall \text{sufficiently regular f}\quad \mathcal{H}(f) = -L^V(f) = -\tfrac{1}{2}\ \textstyle\sum_{1\leq i\leq p}\partial_{x_i}^2 f + V_t\ f \Rightarrow \partial_t q_t = -\mathcal{H}(q_t)\end{aligned} \tag{3.20}$$

In view of (3.18), for any bounded function f we have the Feynman-Kac formula

$$\int f(x)\ q_t(x)\ dx = \mathbb{E}\left(f(X_t^0)\ \exp\left\{-\int_0^t V_s(X_s^0)ds\right\}\right) \simeq_{\epsilon\downarrow 0} \mathbb{E}\left(f(X_{t_n}^{0,\epsilon})\ \prod_{0\leq k<n} G_k\left(X_{t_k}^{0,\epsilon}\right)\right) \tag{3.21}$$

In this context, we also have that

$$-\frac{1}{t}\log \mathbb{P}\left(T>t\right) \simeq_{t\uparrow\infty} E_0 \quad \text{and} \quad \mathbb{P}\left(X_t \in dx \mid T>t\right) \simeq_{t\uparrow\infty} \frac{\varphi_0(x)\ dx}{\int \varphi_0(y)\ dy}$$

where $-E_0$ is the top of the spectrum of the operator L^V on smooth $\mathbb{L}_2$-functions and

$$L^V(\varphi_0) = -E_0\ \varphi_0 \Longleftrightarrow \mathcal{H}(\varphi_0) = E_0\ \varphi_0$$

We refer the reader to section 8.5.2 for a physical derivation of these models, and their extensions to more general diffusions associated with some drift a_t and some diffusion σ_t functions.

We end this section with a brief discussion on the description of the Hamiltonian operator (3.20) associated with a molecule in quantum physics. In this context, a state $x = ((x_{a,i})_{1\leq i\leq N_a}, (x_{e,j})_{1\leq j\leq N_e})$ represents the locations x_a^i of N_a atom nuclei, and the location x_e^j of N_e electrons (we assume that each atom has the same number of electrons) w.r.t. a Cartesian reference frame. The (exact non-relativistic, time-independent molecular) Hamiltonian (3.20) is now given by $\mathcal{H} = -L + \frac{1}{\hbar}V$ with the diffusion generator

$$L := \underbrace{\frac{\hbar}{2}\sum_{1\leq i\leq N_a}\frac{1}{m_{a,j}}\partial^2_{x_{a,i}}}_{:=L^{(a)}\text{ nuclear kinetic energy}} + \underbrace{\frac{\hbar}{2}\sum_{1\leq i\leq N_e}\frac{1}{m_{e,j}}\partial^2_{x_{e,i}}}_{:=L^{(e)}\text{ electronic kinetic energy}}$$

(where $m_{a,j}$ stands for the mass of the j-th nuclei, $m_{e,i}$ stands for the mass of the i-th electron, and $\hbar$ the Planck constant) and the potential function defined in terms of repulsive or attractive Coulomb forces

$$V(x) := \underbrace{\frac{1}{2}\sum_{1\leq i<j\leq N_a}\frac{z_{a,i}z_{a,j}}{\|x_{a,i}-x_{a,j}\|}}_{\text{nuclear repulsion}} + \underbrace{\sum_{1\leq i<j\leq N_e}\frac{e^2}{\|x_{e,i}-x_{e,j}\|}}_{\text{electronic repulsion}} - \underbrace{\frac{1}{2}\sum_{1\leq i\leq N_a}\sum_{1\leq j\leq N_e}\frac{z_{a,i}e^2}{\|x_{a,i}-x_{e,j}\|}}_{\text{electron-nuclear attraction}}$$

for some non negative atomic numbers $z_{a,i}$. The nuclei being much more heavier than electrons (for instance, the proton mass ($1.67\ 10^{-27}$ kg) is 1800 times larger than the electron one ($9.31\ 10^{-31}$ kg)), in the Born-Oppenheimer approximation [62] the nuclei $(x_{a,i})_{1\leq i\leq N_a}$ are fixed parameters, and we reduce the problem to the electronic configuration $x = ((x_{e,j})_{1\leq j\leq N_e})$ associated with the Hamiltonian operator $\mathcal{H} = -L^{(e)} + V$. In physics, the Schrödinger (imaginary time) equation is often written as

$$\hbar\ \partial_t q_t(x) = \hbar\ L(q_t)(x) - V(x)q_t(x)$$

In this situation, the Hamiltonian operator is defined as above by replacing $\mathcal{H}$ by $\hbar\mathcal{H} = -\hbar L + V$. We refer the reader to (8.47) for a physical derivation of these models.

3.5.2 The harmonic oscillator

 In some particular instances, the spectrum of the operator $\mathcal{H} := -L^V$ acting on smooth functions of $\mathbb{L}_2(\mathbb{R}^p)$ can be explicitly computed. For instance, for $p = 1$, $a_t = 0$, $\sigma_t = \sigma := \sqrt{\frac{\hbar^2}{2m}}$ and $V_t(x) = \frac{1}{2}\ m\omega^2\ x^2$, the nonnegative function $q_t(x)$ in (3.19) satisfies the Schrödinger (imaginary time) equation with

$$L^V(f)(x) = \frac{\hbar}{2m}\ \partial_x^2(f)(x) - \frac{1}{2\hbar}\ m\omega^2\ x^2\ f(x)$$

$$\Longleftrightarrow \hbar\ \partial_t q_t(x) = \frac{\hbar^2}{2m}\ \partial_x^2(q_t)(x) - \frac{1}{2}\ m\omega^2\ x^2\ q_t(x)$$

We refer the reader to (8.47) for a physical derivation of these evolution equations. Recalling (3.21) we have the Feynman-Kac representation

$$\int f(x)\ q_t(x)\ dx = \mathbb{E}\left(f(B_t)\ \exp\left\{ -\frac{1}{2\hbar}\ m\omega^2 \int_0^t B_t^2\ ds \right\} \right) \simeq_{\epsilon\downarrow 0} \mathbb{E}\left(f(B^\epsilon_{t_n}) \prod_{0\le k<n} e^{-\frac{\epsilon m\omega^2}{2}\ B^\epsilon_{t_k}} \right) \tag{3.22}$$

with the Gaussian processes

$$dB_t := \sqrt{\frac{\hbar}{m}}\ dW_t \ \simeq_{\epsilon=t_k-t_{k-1}\downarrow 0}\ B^\epsilon_{t_k} - B^\epsilon_{t_{k-1}} := \sqrt{\frac{\hbar\epsilon}{m}}\ \mathcal{W}_k \quad \text{with} \quad \epsilon := (t_k - t_{k-1}) \tag{3.23}$$

In the above display, W_t stands for a standard Brownian motion on the real line, and $\mathcal{W}_k$ a sequence of i.i.d. centered Gaussian r.v. with unit variance. The r.h.s. of the above display is a particular example of the linear-Gaussian filtering models discussed in section 2.2 (cf. (2.6) with $X_n = B^\epsilon_{t_n}$, $\sigma = \sqrt{\frac{\hbar\epsilon}{m}}$, $\tau^2 = \epsilon\ \frac{m\omega^2}{\hbar}$ and $y_n = 0$).

This model is called the harmonic oscillator and it is discussed in section 11.1.2. In this situation, the orthonormal eigenfunctions φ_n associated with the eigenvalues

$$\forall n \ge 0 \qquad E_n = \hbar\ \left(n + \frac{1}{2}\right)\ \omega$$

are defined by

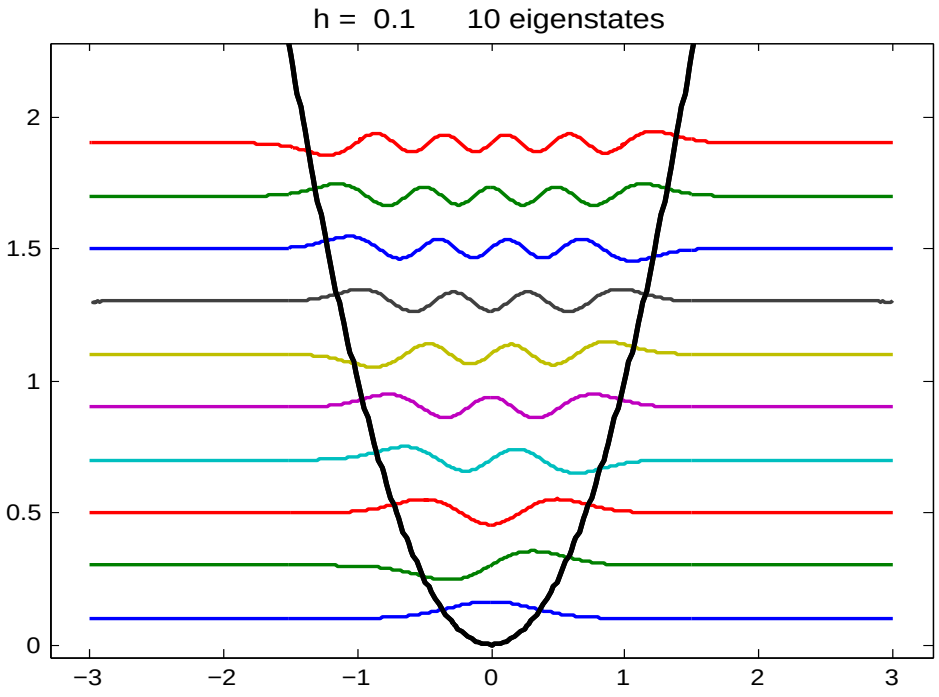
$$\varphi_n(x) := \frac{1}{\sqrt{2^n n! \sqrt{\pi}}}\ \left(\frac{m\omega}{\hbar}\right)^{1/4} \exp\left[-\frac{x^2}{2}\frac{m\omega}{\hbar} \right]\ \mathbb{H}_n\left(\sqrt{\frac{m\omega}{\hbar}} x \right)$$

with the Hermite polynomials $\mathbb{H}_n$ defined by the Rodrigues' formula

$$\mathbb{H}_n(x) = (-1)^n\ e^{x^2}\ \frac{d^n}{dx^n} e^{-x^2} \tag{3.24}$$

The numerical solving of the eigenvalues problem for general potential functions $V(x)$ in 1 dimension can be done using the software `Maltalb Chebfun software` (based on 1-dimensional dynamical system integrations).

The next picture illustrates the first 10 eigenstates associated with $V(x) = x^2$ and $\sigma^2 = 0.1$.



3.5.3 Ground state energies of quantum systems

We consider a time homogeneous Feynman-Kac model

$$Q_t(f)(x) := \mathbb{E}\left(f(X_t)\ \exp\left\{ -\int_0^t V(X_s)ds \right\} \mid X_0 = x \right) \tag{3.25}$$

associated with a $\mathbb{R}^p$ valued Markov process X_t with infinitesimal generator L, and some energy function V. In the above display, the function f stands for a test function (a.k.a. observable in computational physics). For any $s \leq t$, we have the semigroup property

$$\begin{aligned} Q_t(f)(x) &:= \mathbb{E}\left(f(X_t)\ \mathbb{E}\left(\exp\left\{ -\int_0^t V(X_r)dr \right\} \mid X_t \right) \mid X_0 = x \right) \\ &= \mathbb{E}\left(\underbrace{\mathbb{E}\left(f(X_t)\ \exp\left\{ -\int_s^t V(X_r)dr \right\} \mid X_s \right)}_{=Q_{t-s}(f)(X_s)} \exp\left\{ -\int_0^s V(X_r)dr \right\} \mid X_0 = x \right) \\ &= Q_s(Q_{t-s}(f))(x) \end{aligned} \tag{3.26}$$

In addition, in view of (3.25)

$$\begin{aligned} Q_t(f)(x) &:= \mathbb{E}\left(f(X_t)\ \mathbb{E}\left(\exp\left\{ -\int_0^t V(X_r)dr \right\} \mid X_t \right) \mid X_0 = x \right) \\ &= \int f(y)\ \underbrace{\mathbb{E}\left(\exp\left\{ -\int_0^t V(X_s)ds \right\} \mid X_t = y \right)\ \mathbb{P}(X_t \in dy \mid X_0 = x)}_{:=Q_t(x,dy)} \end{aligned} \tag{3.27}$$

When $\mathbb{P}(X_t \in dy \mid X_0 = x)$ has a density $p_t(x,y)$ w.r.t. the Lebesgue measure dy we have

$$Q_t(x,dy) = q_t(x,y)dy \quad \text{with} \quad q_t(x,y) = \mathbb{E}\left(\exp\left\{ -\int_0^t V(X_s)ds \right\} \mid X_t = y \right)\ p_t(x,y) \tag{3.28}$$

We also consider the Feyman-Kac measures

$$\gamma_t(f) = \mathbb{E}\left(f(X_t)\ \exp\left\{ -\int_0^t V(X_s)ds \right\} \right) \quad \text{and} \quad \eta_t(f) = \gamma_t(f)/\gamma_t(1) \tag{3.29}$$

The normalizing constant

$$\mathcal{Z}_t := \gamma_t(1) = \mathbb{E}\left(\exp\left\{ -\int_0^t V(X_s)ds \right\} \right)$$

is also termed the free energy, or the partition function. In section 8.5.2 and section 9.2.3 we will check that for any $s \leq t$

$$\gamma_t = \gamma_s Q_{t-s} \qquad \frac{\partial}{\partial t} Q_t(f) = -\mathcal{H}(Q_t(f)) \quad \text{with the Hamiltonian} \quad \mathcal{H}(f) = -L(f) + Vf$$

and the normalized Schrödinger equation

$$\partial_t \eta_t(f) = \eta_t\left(L(f)\right) + \eta_t(V)\eta_t(f) - \eta_t(fV) = -\eta_t(\mathcal{H}(f)) + \eta_t(V)\eta_t(f) = -\eta_t\left(\left[\mathcal{H} - \eta_t(V)\right](f)\right) \tag{3.30}$$

The operator Q_t is sometimes termed a Feynman-Kac propagator and it is oftenwritten in the exponential form $Q_t = e^{-t\mathcal{H}}$.

A direct calculation shows that

$$\partial_t \log \gamma_t(1) = \frac{1}{\gamma_t(1)} \mathbb{E}\left(\partial_t \exp\left\{-\int_0^t V(X_s)ds\right\}\right) = \frac{\mathbb{E}\left(V(X_t)\ \exp\left\{-\int_0^t V(X_s)ds\right\}\right)}{\mathbb{E}\left(\exp\left\{-\int_0^t V(X_s)ds\right\}\right)} = \eta_t(V)$$

from we conclude that

$$\gamma_t(1) = \exp\left\{-\int_0^t \eta_s(V)\ ds\right\} \Rightarrow -\frac{1}{t}\log \mathcal{Z}_t = -\frac{1}{t}\int_0^t \eta_s(V)\ ds \simeq_{t\uparrow\infty} \eta_\infty(V)$$

for some limiting Feynman-Kac measure η_∞ satisfying the fixed point equation

$$(3.30) \Longrightarrow \eta_\infty(\mathcal{H}(f)) = \eta_\infty(V)\eta_\infty(f) \iff \eta_\infty(L(f)) = \Psi_V(\eta_\infty)(f) \tag{3.31}$$

When L is a self adjoint operator on $\mathbb{L}_2(\mathbb{R}^d)$ (equipped with the Lebesgue measure dx, and the inner product $\langle f_1, f_2\rangle = \int f_1(x)f_2(x)dx$), we have the spectral decomposition

$$Q_t(x, dy) = \sum_{i\geq 0} e^{-tE_i}\ \varphi_i(x)\varphi_i(y)\ dy \tag{3.32}$$

in terms of a sequence of non negative eigenvalues $0 \leq E_0 \leq E_1 \leq \ldots$ and a corresponding set of orthonormal eigenfunctions φ_i, $i \geq 0$, of $\mathcal{H}$; that is, we have that

$$\forall i \geq 0 \qquad \mathcal{H}(\varphi_i) = E_i\varphi_i$$

For a more rigorous and detailed discussion on these spectral decompositions we refer the reader to [185]. The ground state of a quantum mechanical system associated with the Schrödinger Hamiltonian operator $\mathcal{H}$ is its lowest-energy state φ_0. The energy E_0 of the ground state is known as the zero-point energy of the system. An excited state is any state φ_i with energy E_i greater than the one of the ground state.

Notice that

$$\frac{\langle \varphi_0, \mathcal{H}(\varphi_0)\rangle}{\langle \varphi_0, \varphi_0\rangle} = E_0$$

Expressing any normalized function $\varphi \in \mathbb{L}_2(\mathbb{R}^d)$ on the basis of orthonormal eigenfunctions φ_i, $i \geq 0$, we find that

$$\varphi = \sum_{i\geq 0}\langle \varphi, \varphi_i\rangle\ \varphi_i \Rightarrow \begin{array}{rcl} \langle \varphi, \mathcal{H}(\varphi)\rangle & = & \sum_{i\geq 0} E_i\ \langle \varphi, \varphi_i\rangle^2\ \langle \varphi_i, \varphi_i\rangle \\ & = & \sum_{i\geq 0} E_i\ \langle \varphi, \varphi_i\rangle^2 \geq E_0\ \sum_{i\geq 0}\ \langle \varphi, \varphi_i\rangle^2 = E_0 \end{array} \tag{3.33}$$

This yields the variational principle

$$\frac{\langle \varphi, \mathcal{H}(\varphi)\rangle}{\langle \varphi, \varphi\rangle} \geq E_0 \tag{3.34}$$

In addition, we also have that

$$(3.31) \Rightarrow \eta_\infty(\mathcal{H}(\varphi_0)) = E_0\ \eta_\infty(\varphi_0) = \eta_\infty(V)\eta_\infty(\varphi_0) \Rightarrow \eta_\infty(V) = E_0$$

3.5.4 Bra-kets and path integral formalism

In theoretical and computational physics, the state space S is generally the Euclidian space $S = \mathbb{R}^d$. For any absolutely continuous distributions

$$\mu_f(dx) = f(x)\ dx$$

and for functions g on $\mathbb{R}^d$, the Feynman-Kac measures γ_t introduced in (3.29), and the left and right actions of the integral operator Q_t defined in (3.27) on measures and functions are often written in terms of bra-kets

$$\eta_0 = \text{Law}(X_0) = \mu_f \Rightarrow \gamma_t = \eta_0 Q_t = \mu_f Q_t =\prec f|e^{-t\mathcal{H}}$$

as well as

$$Q_t(g) = e^{-t\mathcal{H}}\ |g\succ \quad \text{and} \quad \gamma_t(g) = \eta_0 Q_t(g) = \mu_f Q_t(g) =\prec f|e^{-t\mathcal{H}}|g\succ$$

We further assume that the semigroup $Q_t(x, dy)$ of the Hamiltonian operator

$$\mathcal{H} = -L^V = -L + V$$

has a density $q_t(x, y)$ w.r.t. the Lebesgue measure dy. In this context, the density $q_t(x, y)$ introduced in (3.28) is sometimes written as

$$q_t(x, y) =\prec x|e^{-t\mathcal{H}}|y\succ \quad \text{or} \quad q_t(x, y) =\prec \delta_x|e^{-t\mathcal{H}}|\delta_y\succ$$

so that

$$\int dx\ f(x)\ q_t(x, y)\ g(y)\ dy = \int \ \prec x|e^{-t\mathcal{H}}|y\succ\ f(x)\ g(y)\ dxdy$$

Representing formally functions on the "basis" of delta functions

$$"f(.) = \int\ f(x)\ \delta_x(.)\ dx" \quad \text{in the sense that} \quad \forall y \in \mathbb{R}^d \quad " \int\ f(x)\ \underbrace{\delta_x(y)}_{=1_{x=y}}\ dx = f(y)"$$

using the linearity of the brackets, we arrive at the formal expression

$$\begin{aligned} \prec f|e^{-t\mathcal{H}}|g\succ &= \prec \left(\int\ f(x)\ \delta_x(.)\ dx\right)\ |e^{-t\mathcal{H}}|\ \left(\int\ g(y)\ \delta_y(.)\ dy\right)\ \succ \\ &= \int\ f(x)\ g(y)\ \prec \delta_x|e^{-t\mathcal{H}}|\delta_y\succ\ dxdy = \int\ \prec x|e^{-t\mathcal{H}}|y\succ\ f(x)\ g(y)\ dxdy \end{aligned}$$

In this notation, we have

$$\begin{aligned} \prec f_1|e^{-t\mathcal{H}}|f_2\succ &= \int\ f_1(x)\ \prec x|e^{-t\mathcal{H}}|y\succ\ f_2(y)\ dxdy \\ &= \int\ (\mu_{f_1}Q_t)(dy)\ f_2(y) = \int \mu_{f_1}(dx)\ Q_t(f_2)(x) = \mu_{f_1}Q_t(f_2) \end{aligned}$$

Similarly, the variational principle (3.34) takes the form

$$\frac{\prec \varphi|\mathcal{H}|\varphi\succ}{\prec \varphi, \varphi\succ} \geq E_0 = \frac{\prec \varphi_0|\mathcal{H}|\varphi_0\succ}{\prec \varphi_0, \varphi_0\succ}$$

Dividing $[0, n\Delta t] = [0, t_n]$ into n intervals $([0, \Delta t] \cup \ldots \cup [(n-1)\Delta t, n\Delta t])$ of length Δt, the semigroup property (3.26) implies that

$$\begin{aligned} Q_{n\Delta t}(x_0, dx_n) &= \overbrace{(Q_{\Delta t} \ldots Q_{\Delta t})}^{n \text{ times}}(x_0, dx_n) \\ &= \left[\int\ q_{\Delta t}(x_0, x_1) \ldots q_{\Delta t}(x_{n-1}, x_n)\ dx_1 \ldots dx_{n-1}\right]\ dx_n \\ &= \left[\int\ \left\{\prod_{0\leq k<n} \prec x_k|e^{-\Delta t\ \mathcal{H}}|x_{k+1}\succ\right\}\ dx_1 \ldots dx_{n-1}\right]\ dx_n \end{aligned}$$

Whenever $X_t = B_t$ is the Gaussian process defined in (3.23), and replacing V by $\frac{1}{\hbar}\, V$ in (3.28) we have

$$\prec x_k|e^{-\Delta t\ \mathcal{H}}|x_{k+1}\succ \ \simeq_{\Delta t\downarrow 0}\ e^{-\frac{V(x_k)}{\hbar}\ \Delta t}\ p_{\Delta t}(x_k,x_{k+1}) = \sqrt{\frac{m}{2\pi\hbar\Delta t}}\ e^{-\frac{1}{\hbar}\left[\frac{m}{2}\left(\frac{x_{k+1}-x_k}{\Delta t}\right)^2+V(x_k)\right]\Delta t}$$

In this situation, the discrete time approximation of the integral operator Q_t defined in (3.25) is given by the formula

$$Q_{n\Delta t}(x_0,dx_n) \simeq_{\Delta t\downarrow 0} \left[\int\ \left(\sqrt{\frac{m}{2\pi\hbar\Delta t}}\right)^n\ e^{-\frac{1}{\hbar}\ S_n(x_0,\ldots,x_n)\Delta t}\ dx_1\ldots dx_{n-1}\right]\ dx_n$$

with the so-called Euclidian action functional

$$S_n(x_0,\ldots,x_n) = \sum_{0\leq k<n}\left[\frac{m}{2}\left(\frac{x_{k+1}-x_k}{\Delta t}\right)^2 + V(x_k)\right]$$

Taking formally the limit $\Delta t\downarrow 0$, the density $q_t(x,y)$ is often written in physics literature as a path integral

$$q_t(x,y) = \int_{x_0=x}^{x_t=y}\ \mathcal{D}x\ e^{-\frac{1}{\hbar}\ S_t(x)} \quad \text{with} \quad S_t(x) = \int_0^t \left\{\frac{m}{2}\ \dot{x}_s^2 + V(x_s)\right\}\ ds$$

3.5.5 The φ_0-process Feynman-Kac model

The continuous time version of (2.13) is given by the formula

$$\eta_t(f) := \frac{\gamma_t(f)}{\gamma_t(1)} = \frac{\mathbb{E}\left(\varphi_0^{-1}(X_t^{\varphi_0})\ f(X_t^{\varphi_0})\right)}{\mathbb{E}\left(\varphi_0^{-1}(X_t^{\varphi_0})\right)} = \underbrace{\frac{\langle f,\varphi_0\rangle}{\langle 1,\varphi_0\rangle}}_{:=\eta_\infty(f)} + \mathrm{O}\left(e^{-t(E_1-E_0)}\right) \quad (\Longleftarrow\ (3.32)) \tag{3.35}$$

as soon as $E_1 > E_0$; where $\eta_0 = \mathrm{Law}(X_0)$, and $X_t^{\varphi_0}$ stands for the Markov process with initial distribution

$$\eta_0^{[\varphi_0]} = \Psi_{\varphi_0}(\eta_0)$$

and the infinitesimal generator

$$L^{[\varphi_0]}(f) = L(f) + \varphi_0^{-1}\Gamma_L(\varphi_0,f)$$

with the carré du champ operator

$$\Gamma_L(f,g)(x) := L([f-f(x)][g-g(x)])(x) = L(fg)(x) - f(x)L(g)(x) - g(x)L(f)(x)$$

We also notice that

$$(3.35)\ \Longrightarrow\ \mathbb{P}(X_t^{\varphi_0}\in dx)\ \simeq_{t\uparrow\infty}\ \Psi_{\varphi_0}(\eta_\infty) := \varphi_0^2(x)dx\ /\ \int \varphi_0^2(y)dy \tag{3.36}$$

For instance, for $p=1$ we have

$$L = \frac{1}{2}\partial_x^2 \Rightarrow\ L^{[\varphi_0]}(f)\ =\ \tfrac{1}{2}\ \partial_x^2(f) + (\partial_x\log\varphi_0)\ \partial_x(f) = \tfrac{1}{2}\ \varphi_0^{-2}\ \partial_x\left(\varphi_0^2\ \partial_x(f)\right) \tag{3.37}$$

so that $X_t^{\varphi_0}$ satisfies the Langevin diffusion equation

$$dX_t^{\varphi_0} = (\partial_x\log\varphi_0)\,(X_t^{\varphi_0})\ dt + dW_t$$

where W_t stands for a Brownian motion on the real line. In addition, for any smooth functions (f_1, f_2) with compact support, using a simple integration by part we prove that

$$\begin{aligned}(3.37) \Longrightarrow \int \varphi_0^2(x)\, f_1(x)\, L^{[\varphi_0]}(f_2)(x)\, dx &= \int f_1(x)\, \partial_x\left(\varphi_0^2\, \partial_x(f_2)\right) dx \\ &= -\int \partial_x(f_1)(x)\, \varphi_0^2(x)\, \partial_x(f_2)(x)\, dx \\ &= \int \varphi_0^2(x)\, L^{[\varphi_0]}(f_1)(x)\, f_2(x)\, dx \qquad \text{(by symmetry)}\end{aligned}$$

Using (3.36), this implies that $L^{[\varphi_0]}$ is reversible w.r.t. $\Psi_{\varphi_0}(\eta_\infty)$; that is, we have that

$$(3.36) \implies \Psi_{\varphi_0}(\eta_\infty)\left(f_1\, L^{[\varphi_0]}(f_2)\right) = \Psi_{\varphi_0}(\eta_\infty)\left(L^{[\varphi_0]}(f_1)\, f_2\right)$$

More generally, suppose that L is reversible with respect to some reference non negative measure μ on $\mathbb{R}^p$; that is we have that

$$\mu\left(f_1 L(f_2)\right) = \mu\left(L(f_1) f_2\right) \implies \mu\left(f_1 \mathcal{H}(f_2)\right) = \mu\left(\mathcal{H}(f_1) f_2\right) \tag{3.38}$$

for any couple of smooth functions f_1, f_2. In this situation, the measure η_∞ is given by

$$\eta_\infty = \Psi_{\varphi_0}(\mu)$$

To check this claim, we observe that

$$\mathcal{H}(\varphi_0) = E_0\varphi_0 \quad \Rightarrow \quad \mu(\mathcal{H}(\varphi_0)) = -\overbrace{\mu(L(\varphi_0))}^{=\mu(L(1)\varphi_0)=0} + \mu(V\varphi_0) = E_0\ \mu(\varphi_0) \Rightarrow \Psi_{\varphi_0}(\mu)(V) = E_0$$

This yields that

$$\begin{aligned}\Psi_{\varphi_0}(\mu)(\mathcal{H}(f)) - \Psi_{\varphi_0}(\mu)(V)\Psi_{\varphi_0}(\mu)(f) &= \frac{1}{\mu(\varphi_0)}\ \left[\mu(\varphi_0\mathcal{H}(f)) - \Psi_{\varphi_0}(\mu)(V)\mu(\varphi_0 f)\right] \\ &= \frac{1}{\mu(\varphi_0)}\ \left[\mu(\mathcal{H}(\varphi_0)f) - \Psi_{\varphi_0}(\mu)(V)\mu(\varphi_0 f)\right] \\ &= \Psi_{\varphi_0}(\mu)(f)\ \left[E_0 - \Psi_{\varphi_0}(\mu)(V)\right] = 0\end{aligned}$$

3.5.6 Variational Monte Carlo method

As in the discrete time case discussed in section 2.4, *the ground state φ_0 is usually unknown* and we often use the $\varphi_{\mathcal{T}}$-process $X_t^{\varphi_{\mathcal{T}}}$ associated with a trial energy function (a.k.a. guiding or trial wave function) denoted by $\varphi_{\mathcal{T}}$. In this case, the continuous time version of (2.14) is given by the Feynman-Kac formula

$$\gamma_t(f) = \eta_0(\varphi_{\mathcal{T}})\ \mathbb{E}\left(\varphi_{\mathcal{T}}^{-1}(X_t^{\varphi_{\mathcal{T}}})\ f(X_t^{\varphi_{\mathcal{T}}})\ \exp\left(-\int_0^t V_{\mathcal{T}}(X_s^{\varphi_{\mathcal{T}}})ds\right)\right)$$

with the trial ground state energy (a.k.a. local energy) $V_{\mathcal{T}}$ given by

$$V_{\mathcal{T}} := V - \varphi_{\mathcal{T}}^{-1} L(\varphi_{\mathcal{T}}) = \varphi_{\mathcal{T}}^{-1}\mathcal{H}(\varphi_{\mathcal{T}})$$

In the above display, $X_t^{\varphi_{\mathcal{T}}}$ stands for the $\varphi_{\mathcal{T}}$-twisted process with initial distribution $\eta_0^{[\varphi_{\mathcal{T}}]} = \Psi_{\varphi_{\mathcal{T}}}(\eta_0)$ and infinitesimal generator

$$L_{\mathcal{T}}(f) = L(f) + \varphi_{\mathcal{T}}^{-1}\Gamma_L(\varphi_{\mathcal{T}}, f)$$

We further assume that L is reversible with respect to some non negative measure μ; that is, we have (3.38). In this situation, for any couple of smooth functions f_1 and f_2 we have

$$\begin{aligned}\mu\left(\varphi_{\mathcal{T}}^2\ f_1\ L^{[\varphi_{\mathcal{T}}]}(f_2)\right) &= \mu\left(\varphi_{\mathcal{T}}^2\ f_1\ L(f_2)\right)+\mu\left(\varphi_{\mathcal{T}}^2\ f_1\ \left[\varphi_{\mathcal{T}}^{-1}\Gamma_L(\varphi_{\mathcal{T}},f_2)\right]\right)\\ &= \mu\left(\varphi_{\mathcal{T}}^2\ f_1\ L(f_2)\right)+\mu\left(\varphi_{\mathcal{T}}^2\ f_1\ \left[\varphi_{\mathcal{T}}^{-1}L(\varphi_{\mathcal{T}}f_2)-L(f_2)-\varphi_{\mathcal{T}}^{-1}f_2L(\varphi_{\mathcal{T}})\right]\right)\\ &= \underbrace{\mu\left(\varphi_{\mathcal{T}}\ f_1\ L(\varphi_{\mathcal{T}}f_2)\right)}_{=\mu(\varphi_{\mathcal{T}}\ f_2\ L(\varphi_{\mathcal{T}}f_1))}-\mu\left(\varphi_{\mathcal{T}}\ f_1\ f_2L(\varphi_{\mathcal{T}})\right)=\mu\left(\varphi_{\mathcal{T}}^2\ f_2\ L^{[\varphi_{\mathcal{T}}]}(f_1)\right)\end{aligned}$$

This shows that $L^{[\varphi_{\mathcal{T}}]}$ is reversible w.r.t. $\Psi_{\varphi_{\mathcal{T}}^2}(\mu)$, so that

$$\mathbb{P}(X_t^{\varphi_{\mathcal{T}}}\in dx)\ \simeq_{t\uparrow\infty}\ \mu_{\mathcal{T}}:=\Psi_{\varphi_{\mathcal{T}}^2}(\mu)$$

Arguing as in (3.33), the spectral decomposition (3.32) takes the form

$$Q_t(x,dy)=\sum_{i\geq 0}\ e^{-tE_i}\ \varphi_i(x)\varphi_i(y)\ \mu(dy)$$

and we find that

$$\frac{1}{t}\int_0^t V_{\mathcal{T}}\left(X_s^{\varphi_{\mathcal{T}}}\right)\ ds\simeq_{t\uparrow\infty}\Psi_{\varphi_{\mathcal{T}}^2}(\mu)(V_{\mathcal{T}})=\frac{\mu(\varphi_{\mathcal{T}}^2\ V_{\mathcal{T}})}{\mu(\varphi_{\mathcal{T}}^2)}=\frac{\langle\varphi_{\mathcal{T}},\mathcal{H}(\varphi_{\mathcal{T}})\rangle_\mu}{\langle\varphi,\varphi\rangle_\mu}\geq E_0=\frac{\langle\varphi_0,\mathcal{H}(\varphi_0)\rangle_\mu}{\langle\varphi_0,\varphi_0\rangle_\mu}$$

with the inner product $\langle f_1,f_2\rangle_\mu:=\mu(f_1f_2)$ on $\mathbb{L}_2(\mathbb{R}^d,\mu)$. The above approximation is known as the variational Monte Carlo methods and it only provides an upper bound of the ground state energy.

3.5.7 Twisted models, pilot and trial guiding waves

We return to the Feynman-Kac models discussed in section 3.5.6. We let $\gamma_{\mathcal{T},t}$, $\eta_{\mathcal{T},t}$ and $\eta_{\mathcal{T},\infty}$ be the Feynman-Kac measures defined as γ_t and η_t by replacing (X_t,V) by $(X_t^{\varphi_{\mathcal{T}}},V_{\mathcal{T}})$; that is, we have that

$$\eta_{\mathcal{T},t}(f):=\gamma_{\mathcal{T},t}(f)/\gamma_{\mathcal{T},t}(1)\quad\text{with}\quad\gamma_{\mathcal{T},t}(f)=\mathbb{E}\left(f(X_t^{\varphi_{\mathcal{T}}})\ \exp\left(-\int_0^t V_{\mathcal{T}}(X_s^{\varphi_{\mathcal{T}}})ds\right)\right)\tag{3.39}$$

In this notation, we have

$$\gamma_t(f)=\eta_0(\varphi_{\mathcal{T}})\ \gamma_{\mathcal{T},t}(\varphi_{\mathcal{T}}^{-1}f)\Longleftrightarrow\gamma_{\mathcal{T},t}(f)=\gamma_t(\varphi_{\mathcal{T}}f)/\eta_0(\varphi_{\mathcal{T}})=\eta_t(\varphi_{\mathcal{T}}f)\ \gamma_t(1)/\eta_0(\varphi_{\mathcal{T}})$$

as well as

$$\forall 0\leq t\leq\infty\qquad\eta_t=\Psi_{\varphi_{\mathcal{T}}^{-1}}(\eta_{\mathcal{T},t})\Longleftrightarrow\eta_{\mathcal{T},t}=\Psi_{\varphi_{\mathcal{T}}}(\eta_t)$$

In addition, we have the semigroup evolutions

$$\gamma_{\mathcal{T},t}=\gamma_{\mathcal{T},s}Q_{\mathcal{T},t-s}\quad\text{with}\quad Q_{\mathcal{T},t}(f)(x):=\mathbb{E}\left(f(X_t)\ \exp\left\{-\int_0^t V_{\mathcal{T}}(X_s)ds\right\}\mid X_0=x\right)$$

Arguing as above, we find that

$$\frac{\partial}{\partial t}Q_{\mathcal{T},t}(f)=-\mathcal{H}_{\mathcal{T}}(Q_t(f))\quad\text{with the Hamiltonian}\quad\begin{aligned}\mathcal{H}_{\mathcal{T}}(f)&=-L_{\mathcal{T}}(f)+V_{\mathcal{T}}f\\&=-\varphi_{\mathcal{T}}^{-1}L\left(\varphi_{\mathcal{T}}f\right)-Vf\end{aligned}$$

As underlined in [93], "the role of the trial function φ_T is to guide the stochastic walkers (a.k.a. particles) in the important regions (regions corresponding to an important contribution to the averages)." For a more thorough discussion on these models we refer the reader to section 11.1.3. For a more detailed discussion on the choice of the trial waves functions in quantum systems, we also refer the reader to the review article by M. D. Towler [572].

3.5.8 Non homogeneous models

We return to the particle absorption Feynman-Kac models discussed in the beginning of section 3.5.1. We further suppose that the absorption rate is given by $V_t = \dot{\beta}_t \times V$ (≥ 0), for some $\beta_t \uparrow$ and some energy function V on $\mathbb{R}^p$. Also assume that X_t is given by the non homogeneous Langevin diffusion

$$dX_t = -\beta_t \nabla V(X_t)\ dt + \sqrt{2}\ dW_t \quad \text{with} \quad \text{Law}(X_0) = \mu_{\beta_0} \tag{3.40}$$

In this situation, we have

$$\forall t \geq 0 \qquad \text{Law}(X_t \mid 1_{T>t}) = \mu_{\beta_t} \tag{3.41}$$

and

$$\mathcal{Z}_{\beta_t}/\mathcal{Z}_{\beta_0} = \exp\left\{\int_0^t \dot{\beta}_s\ \mu_{\beta_s}(V) ds\right\}$$

These Feynman-Kac absorption interpretation of the Botzmann-Gibbs measures μ_{β_t} are discussed in section 9.2.4. We emphasize that the particle interpretation of these models are particular instances of the evolutionary type particle models discussed in (2.8). In this context, using a time mesh sequence $t_n - t_{n-1} = \epsilon$, the potential functions

$$G_n(x) = e^{-(\beta_{t_{n+1}} - \beta_{t_n})V(x)} \simeq_{\epsilon \downarrow 0}\ e^{-\dot{\beta}_{t_n} V(x)\epsilon}$$

and M_n are the elementary of a Metropolis Adjusted Langevin Algorithm with fixed inverse temperature β_{t_n}.

3.5.9 Diffusion Monte Carlo models

The imaginary time Schrödinger equation associated with the Hamiltonian operator (3.20), the harmonic oscillator (3.22), the path-integration quantum systems (3.25) and (3.29), the twisted model with pilot and trial guiding waves (3.39), the time inhomogeneous model associated a temperature schedule (3.41), are all particular instances of the Feynman-Kac models associated with some reference Markov process X_t and some possibly non homogeneous potential functions V_t on general state space E.

The discrete time approximation of these path integration formula on a time mesh t_n, with time step $\Delta t = (t_n - t_{n-1})$, is given by

$$\mathbb{E}\left(f(X_{t_n})\ \exp\left\{-\int_0^{t_n} V_s(X_s) ds\right\}\right) \simeq_{\Delta t \downarrow 0}\ \mathbb{E}\left(f(X_{t_n}) \prod_{0 \leq k < n} G_k(X_{t_k})\right) \tag{3.42}$$

with the potential functions

$$G_k(X_{t_k}) = \exp\left(-V_{t_k}(X_{t_k})\ \Delta t\right) \simeq_{\Delta t \downarrow 0}\ \exp\left(-\int_{t_k}^{t_{k+1}} V_s(X_s) ds\right)$$

The Markov chain X_{t_n} can be taken as the values of the process X_t at the time step t_n (when these r.v. are easily sampled), or as an Euler-type discrete time approximation of the process X_t on the time mesh t_n. For instance, if X_t is a jump-diffusion type process (3.2), then we can choose the discrete time approximation scheme presented in (3.4). For the time inhomogeneous Langevin models (3.40) we can also combined at every time step an elementary Euler-type transition with a Metropolis-Hasting adjustment acceptance scheme discussed in section 3.4.2. Finally, for Langevin diffusions X_t on a manifold (3.13) we can choose the Euler-projection transitions (3.14), or an Euler discretization of the Langevin equation on the Riemannian parameter space manifold (3.17).

The r.h.s. discrete time approximation model in (3.42) are particular instances of the Feynman-Kac models discussed in chapter 2. All the branching and mean field particle approximation schemes

discussed in section 2.3 can be used to approximate these path integrals. In computational physics the fixed population branching schemes discussed in section 2.3.2 are also termed Diffusion Monte Carlo methods.

Part II

Linear Monte Carlo methods

Chapter 4

Markov chain models

4.1 Markov chains

4.1.1 Historical, transition and twisted models

Markov chain models are one of the simplest stochastic models developed in probability theory. These processes are characterized by the fact that their past and their future evolutions are independent, given the present value of the chain. In other words, the next state of a Markov chain only depends on its current state.

The methodology, and the stochastic analysis developed in these lecture notes, apply to abstract Markov processes taking values in virtually arbitrary measurable state spaces. This abstract framework encapsulates a surprising wide class of random processes, including Markov chain taking values in transition spaces, as well as excursions, and historical processes taking values in path space models.

For finite and ordered state space valued processes, the evolution of the chain is characterized by a sequence of stochastic matrices. These matrices represent the transitions probability of the chain between two consecutive time integers. In other instances, Markov chains are represented by random dynamical systems, defined in terms of a recursive equation relating the next state, with the current state, and some noisy random variables. In applied probability and engineering sciences, these random dynamical systems are also called nonlinear state space models. Much of probability theory is devoted to the simulation and the analysis of these stochastic processes. Reference books on this subjects are [77, 326, 369, 382, 455, 456, 485, 529, 560]. Concrete examples of Markov chain models can also be found in the books by S. Asmussen, P.W. Glynn [20], O. Cappé, E. Moulines, and T. Rydèn [99], G. Fishman [269], and S.P. Meyn and R.L. Tweedie [455].

In discrete time settings, the random states of these models are defined in terms of a sequence of random variables X_n indexed by the integer time parameter $n \in \mathbb{N}$, and taking values in some measurable state spaces E_n. By the "memoryless" property of the Markov process, the evolution of the random states is characterized by the probability description of the elementary transitions $X_{n-1} \in E_{n-1} \rightsquigarrow X_n \in E_n$ between two consecutive integers; that is, we have that

$$K_n(x_{n-1}, dx_n) = \mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right) \tag{4.1}$$

For instance, the simple Gaussian random walk on $E_n = \mathbb{R}$ starting at the origin is given in terms of a sequence of i.i.d. centered Gaussian r.v. W_n with unit variance by the formulae

$$X_n = X_{n-1} + W_n \Rightarrow K_n(x_{n-1}, dx_n) \propto \exp\left(-\frac{1}{2}(x_n - x_{n-1})^2\right) dx_n \tag{4.2}$$

 The historical process of the Markov chain X_n is defined by the sequence of random paths

$$\mathbf{X_n} = (X_0, \ldots, X_n) \in \mathbf{E_n} := (E_0 \times \ldots \times E_n)$$

In operation research and computer sciences, the historical process is often used to store some the random states of a chain evolving in a solution space, and $\mathbf{X}_n$ is sometimes called the archive. Another model of interest is the Markov chain on transition space defined by

$$\overline{X}_n = (X_n, X_{n+1}) \in \overline{E}_n := (E_n \times E_{n+1})$$

Notice that the Markov transitions of the chain $\mathbf{X}_n$ is given for any

$$\boldsymbol{y_n} = ((y_0, \ldots, y_{n-1}), y_n) = (\boldsymbol{y_{n-1}}, y_n) \in \boldsymbol{E_n} = (\boldsymbol{E_{n-1}} \times E_n)$$

and any $\boldsymbol{x_{n-1}} \in \boldsymbol{E_{n-1}}$ by the following formulae

$$\mathbf{K}_n(\boldsymbol{x_{n-1}}, d\boldsymbol{y_n}) = \delta_{\boldsymbol{x_{n-1}}}(d\boldsymbol{y_{n-1}})\ K_n(y_{n-1}, dx_n) \tag{4.3}$$

 We further assume that Y_n is an auxiliary E_n-valued Markov chain starting at $Y_0 = X_0$ with transitions probabilities $M_n(x_{n-1}, dx_n)$ that dominate $K_n(x_{n-1}, dx_n)$; in the sense that

$$\forall x_{n-1} \in E_{n-1} \qquad K_n(x_{n-1}, dx_n) \ll M_n(x_{n-1}, dx_n)$$

For regular Radon Nikodym derivatives

$$G_n(x_n, x_{n+1}) := \frac{dK_{n+1}(x_n, .)}{dM_{n+1}(x_n, .)}(x_{n+1})$$

we have the Feynman-Kac type change of probability measure

$$\forall n \geq 0 \qquad \mathbb{E}\left(\boldsymbol{f_n}(\overline{\mathbf{X}}_{\boldsymbol{n}})\right) = \mathbb{E}\left(\boldsymbol{f_n}(\overline{\boldsymbol{Y}}_{\boldsymbol{n}}) \prod_{0 \leq k < n} G_k(\overline{Y}_k)\right) \tag{4.4}$$

with the historical processes

$$\overline{\mathbf{X}}_{\boldsymbol{n}} = \left(\overline{X}_0, \ldots, \overline{X}_n\right) \quad \text{and} \quad \overline{\mathbf{Y}}_{\boldsymbol{n}} = \left(\overline{Y}_0, \ldots, \overline{Y}_n\right) \in \overline{\mathbf{E}}_{\mathbf{n}} := \left(\overline{E}_0 \times \ldots \times \overline{E}_n\right)$$

associated with the Markov chains $\overline{X}_n = (X_n, X_{n+1})$ and $\overline{Y}_n = (Y_n, Y_{n+1})$ on transition spaces. The equation (4.4) is valid for any bounded function $\mathbf{f_n}$ on $\overline{\mathbf{E}}_{\mathbf{n}}$. The chain Y_n is sometimes called a twisted Markov chain.

For instance, returning to the example discussed in (4.2) for any $\lambda \in \mathbb{R}$ we can take

$$Y_n = Y_{n-1} + \lambda + W_n \Rightarrow M_n(y_{n-1}, dy_n) \propto \exp\left(-\frac{1}{2}((y_n - y_{n-1}) - \lambda)^2\right) dy_n \tag{4.5}$$

In this situation, we have

$$\begin{aligned} G_n(y_n, y_{n+1}) &:= \exp\left(\frac{1}{2}\left[((y_{n+1} - y_n) - \lambda)^2 - (y_{n+1} - y_n)^2\right]\right) \\ &= \exp\left(\frac{\lambda^2}{2} - \lambda\ (y_{n+1} - y_n)\right) \end{aligned}$$

and therefore

$$\prod_{0\leq k<n} G_k(\overline{Y}_k) = \exp\left(\frac{\lambda^2 n}{2} - \lambda\ Y_n\right) \Rightarrow \mathbb{E}\left(\boldsymbol{f_n}(\overline{\mathbf{X}}_{\boldsymbol{n}})\right) = \mathbb{E}\left(\boldsymbol{f_n}(\overline{\boldsymbol{Y}}_{\boldsymbol{n}})\ \exp\left(-\lambda Y_n + \frac{\lambda^2 n}{2}\right)\right) \tag{4.6}$$

4.1.2 Crude Monte Carlo methodologies

Suppose we are given some integral of the form

$$\eta(f) := \int f(x)\ \eta(dx) = \mathbb{E}(f(X)) \tag{4.7}$$

where $\eta(dx)$ is some probability measure on some state space E, and f some function from E into $\mathbb{R}$ such that $\eta(|f|) = \mathbb{E}(|f(X)|) < \infty$.

The central idea is to sample a sequence of independent random copies $(X^i)_{i\geq 1}$ of the r.v. X and use the so-called empirical average estimates

$$\eta^N(f) := \int f(x)\ \eta^N(dx) = \frac{1}{N}\sum_{1\leq i\leq N} f(X^i) \quad \text{with} \quad \eta^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{X^i} \tag{4.8}$$

In the above display δ_a stands for the Dirac measure at some point $a \in E$.


We set

$$\sqrt{N}\left(\eta^N - \eta\right) := V^N \Longleftrightarrow \eta^N = \eta + \frac{1}{\sqrt{N}}\ V^N \tag{4.9}$$

The r.h.s. formula in the above display can be interpreted as a first order type decomposition of the random deviations between the empirical measure η^N and its limiting value η. In this notation, a simple calculation shows that

$$\mathbb{E}\left(V^N(f)\right) = 0 \quad \text{and} \quad \mathbb{E}\left(V^N(f)^2\right) = \sigma^2(f) \tag{4.10}$$

for any function f s.t. $\eta(f^2) < \infty$ with

$$\begin{aligned} \sigma^2(f) &:= \mathbb{E}\left(f(X)^2\right) - \mathbb{E}\left(f(X)\right)^2 = \eta(f^2) - \eta(f)^2 \\ &= \eta([f-\eta(f)]^2) = \frac{1}{2}\int (f(x)-f(y))^2\ \eta(dx)\eta(dy) \end{aligned} \tag{4.11}$$

Working a little harder we prove that

$$\mathbb{E}(|f(X)|) < \infty \quad \Longrightarrow \quad \lim_{N\to\infty} \eta^N(f) = \eta(f) \quad \mathbb{P}-a.s.$$

We further assume that we have a dedicated Monte Carlo simulation tool to draw independent random samples of the elementary transitions $X_{n-1} \rightsquigarrow X_n$. In this situation, by the law of large numbers, the distribution $\mathbb{P}_n = \mathrm{Law}(\boldsymbol{X_n})$ of the random (historical) trajectories $\boldsymbol{X_n}$ can be approximated by the sequence of occupation measures

$$\mathbb{P}_n^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\boldsymbol{X_n^i}} \quad \text{with} \quad \boldsymbol{X_n^i} := (X_0^i, \ldots, X_n^i)$$

associated with N independent copies $(X_n^i)_{1\leq i\leq N}$ of the stochastic process X_n, as $N \to \infty$. More formally, we have the almost sure convergence

$$\mathbb{P}_n^N(\mathbf{f_n}) \longrightarrow_{N\to\infty} \mathbb{P}_n(\mathbf{f_n}) = \mathbb{E}\left(\mathbf{f_n}(\mathbf{X_n})\right)$$

for any bounded measurable function $\mathbf{f_n}$ on the path space $\boldsymbol{E_n}$. The above convergence estimates can be made precise in various ways using standard tools related to empirical processes associated with i.i.d. random variables.

We return to the Gaussian model discussed in (4.2) and we set

$$\mathbf{f_n}(x_0, \ldots, x_n) = 1_{x_n \geq \lambda n} \quad \text{for some } \lambda > 0$$

In this situation, we have

$$X_n = X_{n-1} + W_n = \sum_{1 \leq k \leq n} W_k \stackrel{\text{in law}}{=} \sqrt{n}\, W_1 \quad \Longrightarrow \quad \mathbb{P}_n(\mathbf{f_n}) = \mathbb{P}\left(W_1 \geq \lambda \sqrt{n}\right) \tag{4.12}$$

Applying Mills inequalities

$$\forall \delta > 0 \quad \frac{1}{\delta + 1/\delta} \frac{1}{\sqrt{2\pi}} e^{-\delta^2/2} \leq \mathbb{P}(W_1 \geq \delta) \leq \frac{1}{\delta} \frac{1}{\sqrt{2\pi}} e^{-\delta a^2/2}$$

to $\delta = \lambda\sqrt{n}$, we have

$$\begin{aligned} & \mathbb{P}_n(\mathbf{f_n}) = \mathbb{P}(X_n \geq \lambda n) \simeq_{n\uparrow\infty} \frac{1}{\lambda} \frac{1}{\sqrt{2\pi n}} e^{-\lambda^2 n/2} \\ \Rightarrow \quad & N\, \mathbb{E}\left([\mathbb{P}_n^N(\mathbf{f_n}) - \mathbb{P}_n(\mathbf{f_n})]^2\right) \quad = \quad \mathbb{P}_n(\mathbf{f_n^2}) - (\mathbb{P}_n(\mathbf{f_n}))^2 \\ & \qquad\qquad\qquad\qquad\quad = \quad \mathbb{P}_n(\mathbf{f_n})\,(1 - \mathbb{P}_n(\mathbf{f_n})) \\ & \qquad\qquad\qquad\qquad\quad \simeq_{n\uparrow\infty} \quad \mathbb{P}_n(\mathbf{f_n}) \simeq_{n\uparrow\infty} \frac{1}{\lambda} \frac{1}{\sqrt{2\pi n}} e^{-\lambda^2 n/2} \end{aligned} \tag{4.13}$$

In some instances, to estimate the expectation (4.7) it is more judicious to use a r.v. Y related to the function f. Suppose that Y has some dominating distribution μ s.t. $\eta \ll \mu$. In this situation, we have

$$g = \frac{d\eta}{d\mu} \Rightarrow \eta(f) := \int f(x)\, \eta(dx) = \int f(y)\, g(y)\, \mu(dy) := \mu(fg) \;\Leftrightarrow\; \mathbb{E}(f(X)) = \mathbb{E}(f(Y)g(Y))$$

In this situation, sampling a sequence of independent random copies $(Y^i)_{i\geq 1}$ of the r.v. Y we have

$$\mu^N(fg) := \int f(y)\, g(y)\, \mu^N(dx) \simeq_{N\uparrow\infty} \mu(fg) \quad \text{with} \quad \mu^N = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{Y^i}$$

The variance of this estimate is now given by

$$N\, \mathbb{E}\left([\mu^N(fg) - \mu(fG)]^2\right) = \mu((fg)^2) - (\mu(fg))^2$$

The measure μ is sometimes called the twisted distribution. For a nonnegative function f, if we choose the Boltzmann-Gibbs measure $\mu = \psi_f(\eta)$ then we have a null variance Monte Carlo sampling scheme:

$$\mu(dy) = \psi_f(\eta)(dy) := \frac{1}{\eta(f)}\, f(y)\, \eta(dy) \Rightarrow f(y)g(y) = \eta(f) \Rightarrow \mu((fg)^2) - (\mu(fg))^2 = 0$$

These optimal twisted distributions are usually untractable. For instance, in the example discussed in (4.12) they coincide with the conditional distributions of the path of a Markov chain X_n given the fact that the terminal point ends at time n above the level λn. Nevertheless, sampling N independent copies $\boldsymbol{Y_n^i}$ of the path $\boldsymbol{Y_n}$ of the twisted chain (4.5), using (4.6) we have

$$\mathbb{P}_n^{\lambda,N}(\mathbf{f_n g_n}) \longrightarrow_{N\to\infty} \mathbb{P}_n^{\lambda}(\mathbf{f_n g_n}) = \mathbb{P}_n(\mathbf{f_n}) = \mathbb{P}(X_n \geq \lambda n) = e^{\frac{\lambda^2 n}{2}}\, \mathbb{E}\left(1_{Y_n \geq \lambda n}\, e^{-\lambda Y_n}\right)$$

with

$$\mathbb{P}_n^\lambda := \mathrm{Law}(\boldsymbol{Y_n}) \simeq_{N\uparrow\infty} \mathbb{P}_n^{\lambda,N} := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\boldsymbol{Y_n^i}}$$

and

$$\mathbf{g_n}(\mathbf{y_n}) = \prod_{0\leq k<n} G_k(y_k, y_{k+1}) = \exp\left(-\lambda y_n + \frac{\lambda^2 n}{2}\right)$$

In this case, the variance of the estimate $\mathbb{P}_n^{\lambda,N}(\mathbf{f_n g_n})$ is given by

$$\begin{aligned} N\,\mathbb{E}\left([\mathbb{P}_n^{\lambda,N}(\mathbf{f_n g_n}) - \mathbb{P}_n^\lambda(\mathbf{f_n g_n})]^2\right) &= \mathbb{P}_n^\lambda((\mathbf{f_n g_n})^2) - (\mathbb{P}_n^\lambda(\mathbf{f_n g_n}))^2 \\ &\simeq_{n\uparrow\infty} \mathbb{P}_n^\lambda((\mathbf{f_n g_n})^2) = \mathbb{E}\left(1_{Y_n\geq\lambda n}\, e^{-\lambda Y_n+\lambda^2 n/2} e^{-\lambda Y_n+\lambda^2 n/2}\right) \\ &= \mathbb{E}\left(1_{X_n\geq\lambda n}\, e^{-\lambda X_n+\lambda^2 n/2}\right) \\ &\leq e^{-\lambda^2 n/2}\,\mathbb{P}\left(X_n\geq\lambda n\right) \\ &= e^{-\lambda^2 n/2}\mathbb{P}_n(\mathbf{f_n}) \simeq_{n\uparrow\infty} \frac{1}{\lambda}\,\frac{1}{\sqrt{2\pi n}}\, e^{-\lambda^2 n} \end{aligned}$$

This shows that the variance of the twisted Markov chain approximation is much smaller that the one (4.13) of the direct Monte Carlo scheme.

4.1.3 Martingale decompositions

We let $\mathcal{F}_n = \sigma(X_p,\ p\leq n)$ be the natural filtration of information generated by the random states X_p of the Markov chain from the origin $p=0$, up to time $p=n$. For any sequence of functions $f_n\in\mathcal{B}(E_n)$ we have

$$f_n(X_n) = f_0(X_0) + \sum_{1\leq p\leq n} \Delta f_p(X_p) \quad \text{with} \quad \Delta f_p(X_p) = (f_p(X_p) - f_{p-1}(X_{p-1}))$$

On the other hand, we have the decomposition

$$\Delta f_p(X_p) = \Delta\mathcal{A}_p(f) + \Delta\mathcal{M}_p(f)$$

with the predictable part

$$\begin{aligned} \Delta\mathcal{A}_p(f) &= \mathbb{E}\left(\Delta f_p(X_p) \mid \mathcal{F}_{p-1}\right) \\ &= \mathbb{E}\left(f_p(X_p) \mid \mathcal{F}_{p-1}\right) - f_{p-1}(X_{p-1}) \\ &= K_p(f_p)(X_{p-1}) - f_{p-1}(X_{p-1}) = [f_p - f_{p-1}](X_{p-1}) + L_p(f_p)(X_{p-1}) \end{aligned}$$

and the martingale part

$$\begin{aligned} \Delta\mathcal{M}_p(f) &= [\Delta f_p(X_p) - \mathbb{E}\left(\Delta f_p(X_p) \mid \mathcal{F}_{p-1}\right)] \\ &= f_p(X_p) - \mathbb{E}\left(f_p(X_p) \mid \mathcal{F}_{p-1}\right) = f_p(X_p) - K_p(f_p)(X_{p-1}) \end{aligned}$$

The terminology "predictable and martingale parts" comes from the fact that the discrete time processes

$$\mathcal{A}_n(f) = \sum_{1\leq p\leq n} \Delta\mathcal{A}_p(f) \quad \text{and} \quad \mathcal{M}_n(f) = \sum_{1\leq p\leq n} \Delta\mathcal{M}_p(f)$$

are predictable, resp. martingales w.r.t. $\mathcal{F}_p$; that is we have that

$$\mathbb{E}\left(\mathcal{A}_n(f) \mid \mathcal{F}_{n-1}\right) = \mathcal{A}_n(f) \quad \text{and} \quad \mathbb{E}\left(\mathcal{M}_n(f) \mid \mathcal{F}_{n-1}\right) = \mathcal{M}_{n-1}(f)$$

We recall that the angle bracket $\langle \mathcal{M} \rangle_n$ (a.k.a. the predictable quadratic variation) of a given martingale $\mathcal{M}_n$ w.r.t. some filtration $\mathcal{F}_n$ is the predictable stochastic process $\langle \mathcal{M} \rangle_n$ s.t. $\mathcal{M}_n^2 - \langle \mathcal{M} \rangle_n$ is a martingale given by

$$\langle \mathcal{M} \rangle_n = \sum_{1 \leq p \leq n} \Delta \langle \mathcal{M} \rangle_p \quad \text{with} \quad \Delta \langle \mathcal{M} \rangle_p := \mathbb{E}\left((\Delta \mathcal{M}_p)^2 \mid \mathcal{F}_{p-1} \right)$$

In our framework, the angle bracket $\langle \mathcal{M}(f) \rangle_n$ of $\mathcal{M}_n(f)$ is given by the local variance increment

$$\begin{aligned} \Delta \langle \mathcal{M}(f) \rangle_p &= \mathbb{E}\left((f_p(X_p) - K_p(f_p)(X_{p-1}))^2 \mid X_{p-1} \right) \\ &= \int K_p(X_{p-1}, dx_p) \; [f_p(x_p) - K_p(f_p)(X_{p-1})]^2 \\ &= \Gamma_{L_p}(f_p, f_p)(X_{p-1}) - (L_p(f_p)(X_{p-1}))^2 \end{aligned}$$

with the "carré du champ" operator

$$\Gamma_{L_p}(f_p, f_p)(x) = L_p((f_p - f_p(x))^2)(x) = L_p(f_p^2)(x) - 2 f_p(x) L_p(f_p)(x)$$

 For any $n \geq 1$ we have

$$\Delta f_n(X_n) = [f_n - f_{n-1}](X_{n-1}) + L_n(f_n)(X_{n-1}) \; + \; \Delta \mathcal{M}_n(f)$$

with a martingale $\mathcal{M}_n(f)$ with predictable angle bracket

$$\Delta \langle \mathcal{M}(f) \rangle_n = \Gamma_{L_n}(f_n, f_n)(X_{n-1}) - (L_n(f_n)(X_{n-1}))^2$$

4.1.4 Evolution semigroups

By construction, the laws of the random states $\eta_n = \text{Law}(X_n)$ of the process satisfy the linear evolution equation

$$\eta_n = \eta_{n-1} K_n \iff \eta_n - \eta_{n-1} = \eta_{n-1} L_n \tag{4.14}$$

with the operators $L_n : \; f \in \mathcal{B}(E_n) \mapsto \mathcal{B}(E_{n-1})$ defined by

$$L_n(f)(x) = (K_n - Id)(f)(x) = \mathbb{E}\left(f(X_n) \mid X_{n-1} = x \right) - f(x)$$

On the other hand, we also have

$$\mathbb{P}\left((X_0, \ldots, X_n) \in d(x_0, \ldots, x_n) \right) = \eta_0(dx_0) \prod_{p=1}^{n} K_p(x_{p-1}, dx_p)$$

Given $X_p = x_p$ for some $p < n$, the law of the random state X_n is given by

$$\mathbb{P}\left(X_n \in dx_n \mid X_p = x_p \right) := K_{p,n}(x_0, dx_p)$$

with the semigroup $K_{p,n}$ of integral operators defined by

$$K_{p,n} = K_{p+1} \ldots K_{n-1} K_n$$

with the convention $K_{n,n} = Id$, the identity operator. In the same way, the distribution η_n of X_n satisfies the integral formula $\eta_n = \eta_p K_{p,n}$.

4.2 Feynman-Kac models

4.2.1 Perturbation semigroups

We let X_n be given reference Markov chain taking values in some state spaces E_n with Markov transitions

$$M_n(x_{n-1}, dx_n) = \mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right)$$

We also consider a collection of nonnegative potential functions G_n on E_n. We let $Q_n(x_{n-1}, dx_n)$ be the positive integral operator from E_{n-1} into E_n defined by

$$Q_n(x_{n-1}, dx_n) = G_{n-1}(x_{n-1})\ M_n(x_{n-1}, dx_n) \tag{4.15}$$

These integral operators can be interpreted as a perturbation of the Markov transitions M_n w.r.t. some potential functions G_n.

We let $Q_{p,n}$, $p \leq n$ be the semigroup defined by

$$Q_{p,n}(x_p, dx_n) \quad = \quad \int_{E_{p+1}\times\ldots\times E_{n-1}} Q_{p+1}(x_p, dx_{p+1}) \ldots Q_n(x_{n-1}, dx_n)$$

By construction, for any function f_n on E_n we have

$$\begin{aligned} Q_{p,n}(f_n)(x_p) \quad &= \quad \int\ Q_{p,n}(x_p, dx_n)\ f_n(x_n) \\ &= \quad \mathbb{E}\left(f_n(X_n) \prod_{p\leq q<n} G_q(X_q) \mid X_p = x_p\right) \end{aligned}$$

This shows that $Q_{p,n}$ is the semigroup of the flow of Feynman-Kac measures γ_n on E_n defined by

$$\gamma_n(f_n) = \mathbb{E}\left(f_n(X_n) \prod_{p\leq q<n} G_q(X_q)\right) = \gamma_p(Q_{p,n}(f_n))$$

for any $p \leq n$, with $\gamma_0 = \eta_0 = \text{Law}(X_0)$. The normalized version of these distributions are given by

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1)$$

4.2.2 Path space measures

We emphasize that γ_n, and resp. η_n, are the marginal w.r.t. the final time horizon n of the measures $\boldsymbol{\gamma_n}$, and resp. $\boldsymbol{\eta_n}$, on the set of paths $\boldsymbol{E_n} = (E_0 \times \ldots \times E_n)$ defined for any measurable function $\boldsymbol{f_n}$ on $\boldsymbol{E_n}$ by

$$\boldsymbol{\gamma_n(f_n)} = \mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{p\leq q<n} \boldsymbol{G_n(X_n)}\right) \quad \text{and} \quad \boldsymbol{\eta_n(f_n)} = \boldsymbol{\gamma_n(f_n)/\gamma_n(1)} \tag{4.16}$$

with

$$\boldsymbol{X_n} = (X_0, \ldots, X_n) \quad \text{and} \quad \boldsymbol{G_n(X_n)} = G_n(X_n)$$

The measures $\boldsymbol{\eta_n} = \mathbb{Q}_n$ are sometimes written in terms of the weighted distribution

$$\mathbb{Q}_n(dx) := \frac{1}{\mathcal{Z}_n}\,\Gamma_n(dx) \quad \text{with} \quad \Gamma_n(dx) := \left\{\prod_{0\le p<n} G_p(x_p)\right\}\,\mathbb{P}_n(dx) \tag{4.17}$$

of the random trajectories of the Markov chain $\mathbb{P}_n = \text{Law}\,(\boldsymbol{X_n})$, with the normalizing constants

$$\mathcal{Z}_n = \gamma_n(1) = \mathbb{E}\left(\prod_{0\le q<n} G_q(X_q)\right)$$

In the above display, $dx = d(x_0,\ldots,x_n)$ stands for an infinitesimal neighborhood of the path $x = (x_0,\ldots,x_n)$.

4.2.3 Partition functions

We underline that the normalizing constants $\mathcal{Z}_n$ (a.k.a. partition functions), as well as the measures $\mathbb{Q}_n$, can be represented in terms of the flow of marginal measures $(\eta_p)_{0\le p\le n}$. More precisely, we have the easily checked multiplicative formulae

$$\mathcal{Z}_n = \mathbb{E}\left(\prod_{0\le p<n} G_p(X_p)\right) = \prod_{0\le p<n} \eta_p(G_p) \tag{4.18}$$

We check this claim using the fact that

$$\mathcal{Z}_{n+1} = \gamma_{n+1}(1) = \gamma_n(G_n) = \frac{\gamma_n(G_n)}{\gamma_n(1)}\;\gamma_n(1) = \eta_n(G)\;\mathcal{Z}_n$$


More generally, using this formula the unnormalized measures $\gamma_n := \mathcal{Z}_n \times \eta_n$ can also be rewritten for any $f_n \in \mathcal{B}_n(f_n)$ in the following form

$$\gamma_n(f_n) = \eta_n(f_n) \times \prod_{0\le p<n} \eta_p(G_p)$$

4.2.4 Backward integration

Next, we present a description of the Feynman-Kac measure $\mathbb{Q}_n$ on path space defined in (4.17) in terms of $(\eta_p)_{0\le p\le n}$. We further assume that the Markov transitions M_n are absolutely continuous with respect to some measures λ_n on E_n, and for any $(x_{n-1}, x_n) \in (E_{n-1}\times E_n)$ we have

$$G_{n-1}(x_{n-1})\;M_n(x_{n-1},dx_n) = H_n(x_{n-1},x_n)\;\lambda_n(dx_n) \tag{4.19}$$

for some density function H_n.

In this situation, for any f on E_{n+1}, and for any $x_n \in E_n$ we have

$$G_n(x_n)M_{n+1}(f)(x_n) = \int\; H_{n+1}(x_n,x_{n+1})\;f(x_{n+1})\;\lambda_{n+1}(dx_{n+1})$$

$$\begin{aligned}\Rightarrow\quad \eta_{n+1}(f) &= \Psi_{G_n}(\eta_n)\,(M_{n+1}(f))\\ &= \int_{E_{n+1}}\;\left[\int_{E_n}\eta_n(dx_n)\;\tfrac{1}{\eta_n(G_n)}\;H_{n+1}(x_n,x_{n+1})\right]\;f(x_{n+1})\;\lambda_{n+1}(dx_{n+1})\end{aligned}$$

This shows that

$$\eta_{n+1}(dx_{n+1}) = \tfrac{1}{\eta_n(G_n)}\ \eta_n\left(H_{n+1}(.,x_{n+1})\right)\ \lambda_{n+1}(dx_{n+1})$$

from which we prove that

$$\begin{aligned}
\mathbb{Q}_n(d(x_0,\ldots,x_n)) &= \frac{1}{\mathcal{Z}_n}\ \eta_0(dx_0)G_0(x_0)\ M_1(x_0,dx_1)\ldots G_{n-1}(x_{n-1})\ M_n(x_{n-1},dx_n)\\
&= \frac{1}{\prod_{0\le p<n}\eta_p(G_p)}\ \eta_0(dx_0)\ \prod_{1\le p\le n} H_p(x_{p-1},x_p)\lambda_p(dx_p)\\
&= \eta_0(dx_0)\ \prod_{1\le p\le n}\frac{H_p(x_{p-1},x_p)}{\eta_{p-1}(H_p(.,x_p))}\ \eta_p(dx_p)
\end{aligned}$$

 This implies that

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) = \eta_n(dx_n)\ \prod_{q=1}^{n}\mathbb{M}_{q,\eta_{q-1}}(x_q,dx_{q-1}) \tag{4.20}$$

with the collection of Markov transitions

$$\mathbb{M}_{n+1,\eta_n}(x,dy)\ =\ \frac{\eta_n(dy)\ H_{n+1}(y,x)}{\eta_n\left(H_{n+1}(.,x)\right)}$$

If we take the unit potential functions $G_n = 1$, the backward formula (4.20) reduces to the conventional backward representation of conditional distribution of the random paths $(X_0,\ldots,X_{n-1})$ given the terminal time X_n; that is we have that

$$\mathbb{P}\left((X_0,X_1,\ldots,X_{n-1})\in d(x_0,x_1,\ldots,x_{n-1})\ \middle|\ X_n=x_n\right)$$

$$= \mathbb{M}_{n,\eta_{n-1}}(x_n,dx_{n-1})\cdots\mathbb{M}_{2,\eta_1}(x_2,dx_1)\mathbb{M}_{1,\eta_0}(x_1,dx_0)$$

To the best of our knowledge, these forward-backward representations of Feynman-Kac measures were introduced by Ruslan L. Stratonovitch in the early 1960s in the context of nonlinear filtering [554]. For a more thorough discussion on these backward Markov chain models, and their application in advanced signal processing, and hidden Markov chain problems, we also refer the reader to [232] and to a series of articles of the author with A. Doucet and S.S. Singh [203, 204, 205].

4.2.5 Spatial branching processes

We denote by $\mathbf{E} = \cup_{p\ge 0}E^p$ the state space of a branching process with individuals taking values on some measurable state space E. The integer $p \ge 0$ represents the size of the population. For $p = 0$ we use the convention $E^0 := \{c\}$, where c stands for an auxiliary cemetery state.

We let $M_n(x,dy)$, with $n \ge 1$, be a sequence of Markov transitions from E into inself, and by $(g_n^i(x))_{i\ge 1, x\in E, n\ge 0}$ we denote a collection of integer number-valued random variables with uniformly finite first moments. We further assume that for any $x \in E$, and any $n \ge 0$, $(g_n^i(x))_{i\ge 1}$ are identically distributed, and we set

$$G_n(x) := \mathbb{E}(g_n^i(x))$$

Our branching process is defined as follows. We start at some point x_0 with a single particle, that is $p_0 = 1$ and $\zeta_0 = \zeta_0^1 = x_0 \in E^{p_0} = E$. This particle branches into $\widehat{p}_0$ offsprings $\widehat{\zeta}_0 = (\widehat{\zeta}_0^1,\ldots,\widehat{\zeta}_0^{\widehat{p}_0}) \in E^{\widehat{p}_0}$, with $\widehat{p}_0 = g_0^1(\zeta_0^1)$.

Each of these individuals explores randomly the state space E, according to the transition M_1. At the end of this mutation step, we have a population of $p_1 = \widehat{p}_0$ particles $\zeta_1^i \in E$ with distribution

$M_1(\widehat{\zeta}_0^i, .)$, $i = 1, \ldots, p_1$. Then each of these particles ζ_1^i branches into $g_1^1(\zeta_1^i)$ offsprings. At the end of this transition, we have $\widehat{p}_1$ particles $\widehat{\zeta}_1 = (\widehat{\zeta}_1^1, \ldots, \widehat{\zeta}_1^{\widehat{p}_1}) \in E_1^{\widehat{p}_1}$, with $\widehat{p}_1 = \sum_{i=1}^{p_1} g_1^i(\zeta_1^i)$.

Then, each of these individuals explores randomly the state space E, according to the transition M_2, and so on.

Whenever the system dies, $\widehat{p}_n = 0$ at a given time n, we set $\widehat{\zeta}_q = \zeta_{q+1} = c$, and $\widehat{p}_q = p_{q+1} = 0$, for any $q \geq n$.

By construction, we have $p_{n+1} = \widehat{p}_n$, and $\sum_{i=1}^{\widehat{p}_n} f(\widehat{\zeta}_n^i) = \sum_{i=1}^{p_n} g_n^i(\zeta_n^i)\ f(\zeta_n^i)$, for any function $f \in \mathcal{B}_b(E)$. If we consider the random measures

$$\mathcal{X}_n = \sum_{i=1}^{p_n} \delta_{\zeta_n^i} \quad \text{and} \quad \widehat{\mathcal{X}}_n = \sum_{i=1}^{\widehat{p}_n} \delta_{\widehat{\zeta}_n^i}$$

then, we find that

$$\mathbb{E}(\widehat{\mathcal{X}}_n(f) \mid \zeta_n) = \mathcal{X}_n(G_n\ f) \quad \text{and} \quad \mathbb{E}(\mathcal{X}_{n+1}(f) \mid \widehat{\zeta}_n) = \widehat{\mathcal{X}}_n(M_n(f))$$

This clearly implies that

$$\mathbb{E}(\mathcal{X}_{n+1}(f) \mid \zeta_n) = \mathcal{X}_n(G_n M_{n+1}(f))$$

We readily conclude that the first moments of the branching distributions $\mathcal{X}_n$ are given by the Feynman-Kac model

$$\mathbb{E}(\mathcal{X}_n(f)) = \mathbb{E}_{x_0}\left(f(X_n) \prod_{0 \leq k < n} G_k(X_k)\right) := \gamma_n(f)$$

In the above display, X_n stands for the Markov chain on E with Markov transitions M_n. In this interpretation, the mean number of individuals in the current population is given by $\mathbb{E}(\mathcal{X}_n(1)) = \gamma_n(1)$.

In probability theory, the stochastic process $\mathcal{X}_n$ is called a Branching Markov chain. The long time behavior of these branching models, their connections with particle absorption models, and their applications in physics and biology is a rapidly developing subject in probability theory. We refer the reader to the series of articles [19, 20, 21, 54, 76, 358, 375, 376, 465], the more recent studies [5, 37, 51, 212, 214, 215, 337, 536], and references therein.

4.3 Sub-Markov models

4.3.1 Particle absorption models

In probability theory, particle absorption models are represented by Markov chains evolving in a deterministic, or in a random, environment associated with some absorption rate functions.

The interpretation of the absorption event clearly depends on their application models. In optical ray propagation problems, the event of interest is related to photon absorptions [500]. In particle physics or in chemistry, the absorption rate is dictated by the energy of an electronic or macro-molecular configuration. For a more detailed discussion on these models, and their applications to the computation of Schrödinger ground state energies, we refer the reader to the series of articles [97, 170, 184, 185, 516]. In natural evolution theory, as well as in population model analysis, the absorption event is often related to an extinction probability. Further details on these applications can be found in the series of articles [437, 449, 583, 584, 585].

Absorption and critical type events can also be thought of as network overflows in complex queueing systems [252] and production systems [495]. Absorbed Markov chain also used in Web engineering [496] and bio-chemistry [414], as well as in environmental analysis [598], and in many others scientific disciplines.

This rather extraordinary variety of application domains is not really surprising, since all of these absorption models can be represented by a Feynman-Kac model. Inversely, we emphasize that *any* Feynman-Kac model (4.17) can be interpreted as the distribution of the random trajectories of a Markov chain evolving in an absorbing environment.

We consider a collection of measurable state spaces E_n and an auxiliary coffin, or cemetery, state c. We set $E_{n,c} = E_n \cup \{c\}$. We also denote by G_n some $[0,1]$-valued potential functions on E_n, and M_{n+1} some Markov transitions from E_n, into E_{n+1}. We define an $E_{n,c}$-valued Markov chain X_n^c with two separate killing/exploration transitions:

$$X_n^c \xrightarrow{\text{killing}} \widehat{X}_n^c \xrightarrow{\text{exploration}} X_{n+1}^c \tag{4.21}$$

These killing/exploration mechanisms are defined as follows:

- **Killing:** If $X_n^c = c$, we set $\widehat{X}_n^c = c$. Otherwise the particle X_n^c is still alive. In this case, with a probability $G_n(X_n^c)$, it remains in the same site, so that $\widehat{X}_n^c = X_n^c$; and with a probability $1 - G_n(X_n^c)$, it is killed, and we set $\widehat{X}_n^c = c$.

- **Exploration:** Once a particle has been killed, it cannot be brought back to life; so if $\widehat{X}_n^c = c$, then we set $\widehat{X}_p^c = X_p = c$, for any $p > n$. Otherwise, the particle $\widehat{X}_n^c \in E_n$ evolves to a new location X_{n+1}^c in E_{n+1}, randomly chosen according to the distribution $M_{n+1}(X_n^c, dx_{n+1})$.

Definition 4.3.1 *The Markov chain X_n^c defined above is called a Markov chain with the absorption rates $(1 - G_n)$, and the free exploration transitions M_n, on the state spaces E_n.*

Notice that the Markov chain X_n^c on the augmented state spaces $E_{n,c}$ can be interpreted as a conventional Markov chain, with a single absorbing state $\{c\}$, as soon as $M_n(x_n, \{x_n\}) \neq 1$ for any $x_n \in E_n$. Inversely, any Markov chain with a single absorbing state can be represented in this form.

In branching processes and population dynamics literature, the model X_n^c often represents the number of individuals of a given species [268, 305, 552]. Each individual can die or reproduce. The state $0 \in E_n = \mathbb{N}$ is interpreted as a trap, or as a hard obstacle, in the sense that the species disappears as soon as X_n^c hits 0. For a more thorough discussion on particle motions in an absorbing medium with hard and soft obstacles and their application domains, we refer the reader to the series of articles [185, 184, 516], the monograph [172], and the more recent lecture notes [195].

We end this section with a Feynman-Kac formulation of particle absorption models. We denote by X_n the Markov chain on E_n, with elementary transitions M_n. In this notation, the Feynman-Kac measures $\mathbb{Q}_n$ associated with the parameters (G_n, M_n), and defined in (4.17), represent the conditional distributions of the random paths of a nonabsorbed Markov particle. To see this claim, we let T be the killing time; that is, the first time at which the particle enters in the cemetery state

$$T = \inf \{n \geq 0 \; ; \; \widehat{X}_n^c = c\}$$

By construction, we have

$$\begin{aligned} \mathbb{P}(T \geq n) &= \mathbb{P}(\widehat{X}_0^c \in E_0, \ldots, \widehat{X}_{n-1}^c \in E_{n-1}) \\ &= \int_{E_0 \times \ldots \times E_{n-1}} \eta_0(dx_0)\; G_0(x_0) \prod_{1 \leq p < n} (M_p(x_{p-1}, dx_p) G_p(x_p)) \end{aligned}$$

This shows that the normalizing constants $\mathcal{Z}_n$ of the Feynman-Kac measures $\mathbb{Q}_n$ represent the probability for the particle to be alive at time $n-1$; that is, we have that

$$\mathcal{Z}_n = \mathbb{P}(T \geq n) = \mathbb{E}\left(\prod_{0\leq p<n} G_p(X_p)\right)$$

In the above display, X_n stands for a Markov chain on E_n, with initial distribution η_0 and elementary Markov transitions M_n.

In the same vein, in terms of the n-th time marginal Feynman-Kac models we have

$$\mathbb{E}(f(X_n^c)\ 1_{T\geq n}) = \gamma_n(f_n) := \mathbb{E}\left[f_n(X_n)\left\{\prod_{0\leq p<n} G(X_p)\right\}\right] \tag{4.22}$$

$$\mathbb{E}(f(X_n^c) \mid T \geq n) = \eta_n(f_n) := \gamma_n(f_n)/\gamma_n(1) \tag{4.23}$$

Using these formulae, we also find that

$$\mathcal{Z}_n = \mathbb{P}(T \geq n) = \gamma_n(1) = \mathbb{E}\left(\prod_{0\leq p<n} G_p(X_p)\right) = \prod_{0\leq p<n} \eta_p(G_p)$$


More generally, similar arguments yield that is the distribution of a particle conditional upon being alive at time $n-1$ that is defined by the Feynman-Kac model introduced in (4.17); that is, we have that

$$\mathbb{Q}_n(d(x_0,\ldots,x_n)) = \mathbb{P}\left((X_0^c,\ldots,X_n^c) \in d(x_0,\ldots,x_n) \ \mid T \geq n\right)$$

Inversely, any Feynman-Kac model of the form (4.17) associated with some bounded potential functions G_n can be interpreted in terms of a particle absorption model. To prove this claim, we further assume that $\|G_n\| \leq c_n$ for some finite constant $c_n < \infty$. We let X_n^c be the Markov chain on $E_{n,c}$ defined in (4.21) with absorption rate $(1 - G_n(x_n)/c_n)$. By construction, we readily check that

$$\mathbb{Q}_n := \mathrm{Law}\left((X_0^c,\ldots,X_n^c) \ \mid T \geq n\right)$$

We end this section with a couple of particle absorption model with hard obstacles:

• If we choose indicator potential functions $G_n = 1_{A_n}$, then we have

$$\begin{aligned}\mathcal{Z}_n &= \mathbb{P}\left(X_p \in A_p\ ,\ \forall 0 \leq p < n\right)\\ \mathbb{Q}_n &= \mathrm{Law}\left((X_0,\ldots,X_n) \mid X_p \in A_p\ ,\ \forall 0 \leq p < n\right)\end{aligned}$$

• We assume that $X_n = (X'_0,\ldots,X'_n)$ is the historical process associated with a random walk evolving in a d-dimensional lattice $E = \mathbb{Z}^d$. In this situation, if we set $G_n(X_n) = 1_{\mathbb{Z}^d - \{X'_0,\ldots,X'_{n-1}\}}$, then we find that

$$\begin{aligned}\mathcal{Z}_n &= \mathbb{P}\left(X'_p \neq X'_q\ ,\ \forall 0 \leq p < q < n\right)\\ \mathbb{Q}_n &= \mathrm{Law}\left((X'_0,\ldots,X'_n) \mid X'_p \neq X'_q\ ,\ \forall 0 \leq p < q < n\right)\end{aligned}$$

4.3.2 Doob h-processes

We consider the time homogeneous Feynman-Kac model $(\Gamma_n, \mathbb{Q}_n)$, associated with the parameters $(E_n, G_n, M_n) = (E, G, M)$ on some measurable state space E, defined in (4.17). We also set

$$Q(x, dy) := G(x)M(x, dy)$$

We also assume that G is uniformly bounded above and below by some positive constant, and the Markov transition M is reversible w.r.t. some probability measure μ on E, with $M(x, .) \simeq \mu$ and $dM(x, .)/d\mu \in \mathbb{L}_2(\mu)$. We denote by λ the largest eigenvalue of the integral operator Q on $\mathbb{L}_2(\mu)$, and by $h(x)$ a positive eigenvector

$$Q(h) = \lambda h$$

Under some regularity conditions on (G, M), there exists some constant $\rho \geq 1$ such that

$$1/\rho \leq h(x)/h(y) \leq \rho \tag{4.24}$$

for any $x, y \in E$. For instance, let us suppose that

$$M(x, dz) \geq \epsilon\, M(y, dz) \quad \text{and} \quad G(x) \leq g\, G(y)$$

for some $\epsilon \in]0, 1]$ and some $g < \infty$. In this situation, we have

$$Q(h)(x)/Q(h)(y) = h(x)/h(y) \leq \rho \quad \text{with} \quad \rho \leq g/\epsilon$$

The Doob h-process, corresponding to the ground state eigenfunction h defined above, is a Markov chain X_n^h, with initial distribution $\eta_0^h = \Psi_h(\eta_0)$, and the Markov transition

$$M^h(x, dy) := \frac{1}{\lambda} \times h^{-1}(x)Q(x, dy)h(y) = \frac{M(x, dy)h(y)}{M(h)(x)}$$

We also denote by η_n^h the distribution of the random state X_n^h starting with initial distribution η_0^h; that is, we have that

$$\text{Law}(X_n^h) = \eta_n^h = \eta_0^h (M^h)^n$$

Our next objective is to connect the distribution of the paths of the h-process

$$\mathbb{P}\left((X_0^h, \ldots, X_n^h) \in d(x_0, \ldots, x_n)\right) = \eta_0^h(dx_0)M^h(x_0, dx_1) \ldots M^h(x_{n-1}, dx_n)$$

with the Feynman-Kac measures Γ_n and $\mathbb{Q}_n$ introduced in (4.17). Firstly, by construction we have

$$G = \lambda \times h/M(h)$$

and therefore

$$\begin{aligned}
\Gamma_n(d(x_0, \ldots, x_n)) &= \eta_0(dx_0) \left\{ \prod_{0 \leq p < n} G(x_p) \right\} \times \left\{ \prod_{1 \leq p \leq n} M(x_{p-1}, dx_p) \right\} \\
&= \lambda^n\, \eta_0(dx_0)\; h(x_0) \left\{ \prod_{1 \leq p \leq n} \frac{M(x_{p-1}, dx_p)h(x_p)}{M(h)(x_{p-1})} \right\} \frac{1}{h(x_n)}
\end{aligned}$$

We conclude that

$$\Gamma_n(d(x_0, \ldots, x_n)) = \lambda^n\; \eta_0(h)\; \mathbb{P}_n^h(d(x_0, \ldots, x_n))\; \frac{1}{h(x_n)}$$

where $\mathbb{P}_n^h$ stands for the law of the historical process

$$\mathbf{X_n^h} = (X_0^h, \ldots, X_n^h)$$

This clearly implies that

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \frac{1}{\mathbb{E}(h^{-1}(X_n^h))}\ h^{-1}(x_n)\ \ \mathbb{P}_n^h(d(x_0, \ldots, x_n))$$

with the normalizing constants

$$\mathcal{Z}_n = \lambda^n\ \eta_0(h)\ \mathbb{E}(h^{-1}(X_n^h))$$

4.3.3 Quasi-invariant measures

Under condition (4.24), using the multiplicative formula (4.18) we also have that

$$\frac{1}{n}\log \mathcal{Z}_n = \frac{1}{n}\sum_{0\leq p<n} \log \eta_p(G) = \log\lambda + \frac{1}{n}\log\Big(\eta_0(h)\ \mathbb{E}(h^{-1}(X_n^h))\Big)$$

and therefore

$$\log\lambda - \frac{1}{n}\log\rho \leq \frac{1}{n}\log \mathcal{Z}_n = \frac{1}{n}\sum_{0\leq p<n} \log \eta_p(G) \leq \log\lambda + \frac{1}{n}\log\rho \tag{4.25}$$

from which we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{0\leq p<n} \log \eta_p(G) = \log\lambda$$

In terms of the h-process, the n-th time marginal γ_n of the Feynman-Kac measures Γ_n takes the following form

$$\gamma_n(f) = \lambda^n\ \eta_0(h)\ \eta_0^h(M^h)^n(f/h) = \lambda^n\ \eta_0(h)\ \eta_n^h(f/h)$$

In terms of particle absorption models we have

$$\mathrm{Law}((X_0^c, \ldots, X_n^c)\mid T^c \geq n) = \frac{1}{\mathbb{E}(h^{-1}(X_n^h))}\ h^{-1}(X_n^h)\ \ d\mathbb{P}_n^h$$

and

$$\mathcal{Z}_n = \mathbb{P}\left(T^c \geq n\right) = \lambda^n\ \eta_0(h)\ \mathbb{E}(h^{-1}(X_n^h)) \longrightarrow_{n\uparrow\infty} 0 \tag{4.26}$$

Whenever it exists, the Yaglom limit of the measure η_0 is defined as the limiting of measure

$$\eta_n \longrightarrow_{n\uparrow\infty} \eta_\infty \tag{4.27}$$

of the Feynman-Kac flow η_n, when n tends to infinity. We also say that η_0 is a quasi-invariant measure as we have $\eta_0 = \eta_n$, for any time step. When the Feynman-Kac flow η_n is asymptotically stable, in the sense that it forgets its initial conditions, we also say that the quasi-invariant measure η_∞ is the Yaglom measure.

Quantitative convergence estimates of the limiting formulae (4.27) can be derived using the stability properties of the Feynman-Kac models. For a more thorough discussion on these particle absorption models, we refer the reader to the series of articles of the author with A. Guionnet [178, 179], L. Miclo [170, 185] and A. Doucet [184], as well as the monograph [172].

4.4 Stability properties

4.4.1 Dobrushin ergodic coefficient

This section provides a very brief treatise on the regularity properties of Markov semigroups. For a more thorough discussion, we refer the reader to [455]. We recall that the total variation distance $\|\mu_1 - \mu_2\|_{\rm tv}$ between probability measures μ_1, $\mu_2 \in \mathcal{P}(E)$ is defined by

$$\begin{aligned} \|\mu_1 - \mu_2\|_{\rm tv} &= \sup_{A\in\mathcal{E}} |\mu_1(A) - \mu_2(A)| \\ &= 2^{-1} \sup\{|\mu_1(f) - \mu_2(f)|\,; \quad \|f\| \le 1\} = \sup\{|\mu_1(f) - \mu_2(f)|\,; \quad {\rm osc}(f) \le 1\} \end{aligned}$$

A probability measure π on some state space E is invariant w.r.t. some time homogeneous Markov transition K if we have $\pi K = \pi$. The measure π is reversible w.r.t. K if we have

$$\pi(dx)\ K(x,dy) = \pi(dy)\ K(y,dx)$$

or equivalently, for any pair of functions $(f,g) \in \mathcal{B}(E)^2$

$$\pi(fK(g)) = \pi(K(f)g)$$

We recall that the Dobrushin ergodic coefficient $\beta(K)$ of a Markov transition K from E_1 into E_2, is the norm of the operator K from $\mathcal{M}_0(E_1)$ into $\mathcal{M}_0(E_2)$; that is, we have the equivalent formulations

$$\begin{aligned} \beta(K) &= \sup \|K(x,.) - K(y,.)\|_{\rm tv} = \sup_{\mu\in\mathcal{M}_0(E_1)} \|\mu K\|_{\rm tv}/\|\mu\|_{\rm tv} \\ &= \sup\{{\rm osc}(K(f))\ :\ {\rm osc}(f) \le 1\} \end{aligned}$$

with the first supremum taken over all $(x,y) \in E_1^2$.

We also have the contraction inequalities

$$\|\mu K - \nu K\|_{\rm tv} \le \beta(K)\ \|\mu - \nu\|_{\rm tv} \quad \text{and} \quad {\rm osc}(K(f)) \le \beta(K)\ {\rm osc}(f) \tag{4.28}$$

We check these assertions using the fact that

$$\begin{aligned} \|\mu_1 M - \mu_2 M\|_{tv} &= \sup_{f\ :{\rm osc}(f)\le 1} \left({\rm osc}(M(f)) \times \left| (\mu_1 - \mu_2)\left[\frac{M(f)}{{\rm osc}(M(f))} \right] \right| \right) \\ &= \beta(M) \times \sup_{g\ :{\rm osc}(g)\le 1} |(\mu_1 - \mu_2)(g)| \\ &= \beta(M) \times \|\mu_1 - \mu_2\|_{tv} \end{aligned}$$

for any Markov transition M.

Several rather crude estimates can be underlined. For instance, we have

$$(\forall x,y,z \in E_1 \quad K(x,dz) \ge \epsilon\ K(y,dz)) \Rightarrow \beta(K) \le (1-\epsilon)$$

In the same vein, we have $\beta(K) \le (1-\epsilon)$ as soon as

$$\forall x,y \in E_1 \quad K(x,dy) \ge \epsilon\ \mu(dy) \quad \text{for some} \quad \mu \in \mathcal{P}(E_2) \tag{4.29}$$

The last assertion comes from the fact that the Markov transition

$$K_\mu(x,dy) = \frac{1}{1-\epsilon}\,[K(x,dy) - \epsilon\mu(dy)]$$

is such that

$$[K(x,dz) - K(y,dz)] = (1-\epsilon)\ [K_\mu(x,dz) - K_\mu(y,dz)]$$

Given a pair of Markov transitions K_1 from E_1 into E_2, and K_2 from E_2 into E_3 , and any $f_3 \in \mathrm{Osc}(E_3)$ we have

$$\mathrm{osc}(K_1K_2(f_3)) \leq \beta(K_2)\ \mathrm{osc}(K_1(f_2)) \leq \beta(K_1)\beta(K_2)\ \mathrm{osc}(f_1)$$

with

$$f_2 = K_2(f_3)/\beta(K_2) \in \mathrm{Osc}(E_2) \quad \text{and} \quad f_1 = K_1(f_2)/\beta(K_1) \in \mathrm{Osc}(E_1)$$

This clearly implies that $\beta(K_1K_2) \leq \beta(K_1)\beta(K_2)$. Iterating this argument, for any collection of Markov transitions K_n such that $\beta(K_n) \leq (1-\epsilon)$ we have the quantitative contraction estimate

$$\beta(K_1K_2\dots K_n) \leq \prod_{1\leq p\leq n} \beta(K_p) \leq (1-\epsilon)^n$$

In this situation, in time homogeneous settings we have

$$\|\mu_1K^n - \mu_2K^n\|_{tv} \leq (1-\epsilon)^n\ \|\mu_1 - \mu_2\|_{tv}\ \to_{n\to\infty}\ 0 \tag{4.30}$$

from which we conclude that there exists a unique measure π such that $\pi = \pi K$.

4.4.2 Boltzmann-Gibbs transformations

Given a positive and bounded potential function G on E, we denote by Ψ_G the Boltzmann-Gibbs mapping from $\mathcal{P}(E)$ into itself, defined for any $\mu \in \mathcal{P}(E)$ by

$$\Psi_G(\mu)(dx) = \frac{1}{\mu(G)}\ G(x)\ \mu(dx) \tag{4.31}$$

There is no loss of generality to assume that G is a $]0,1]$-valued function. For $[0,1]$-valued potential functions, the transformation is only defined on measures μ s.t. $\mu(G) > 0$. To avoid unnecessary repetition of technical abstract conditions, and unless otherwise stated, in the further development of this section we frame the standing assumption that G is chosen so that

$$g := \sup_{x,y} (G(x)/G(y)) < \infty \tag{4.32}$$

Next we present a rather strong Lipschitz type estimate. For any pair of measures μ and ν, and any bounded positive function G, we have

$$\|\Psi_G(\mu) - \Psi_G(\nu)\|_{\mathrm{tv}} \leq \frac{\|G\|}{\mu(G)\vee\nu(G)}\ \|\mu-\nu\|_{\mathrm{tv}} \tag{4.33}$$

There is no loss of generality to assume that G is a $]0,1]$-valued function. We prove (4.33) using the fact that the mapping Ψ_G can be expressed in the following form

$$\Psi_G(\mu) = \mu S_\mu \tag{4.34}$$

with the Markov transitions

$$S_\mu(x,dy) = G(x)\ \delta_x(dy) + (1-G(x))\ \ \Psi_G(\mu)(dy) \tag{4.35}$$

On the other hand, we notice that

$$\begin{aligned} \Psi_G(\mu) - \Psi_G(\nu) &= (\mu-\nu)S_\mu + \nu(S_\mu - S_\nu) \\ \nu(S_\mu - S_\nu) &= (1-\nu(G))\,[\Psi_G(\mu) - \Psi_G(\nu)] \end{aligned}$$

from which we find the formula

$$\Psi_G(\mu) - \Psi_G(\nu) = \frac{1}{\nu(G)}\ (\mu - \nu)S_\mu \tag{4.36}$$

In addition, we have

$$S_\mu(x, .) \geq (1 - \|G\|)\ \Psi_G(\mu) \implies \beta(S_\mu) \leq \|G\|$$

The end of the proof of (4.33) is now clear. This ends the proof of (4.33).

We end this section with an interesting contraction property of a Markov transition

$$M_G(x, dy) = \frac{M(x, dy)G(y)}{M(G)(x)} = \Psi_G(\delta_x M)(dy) \tag{4.37}$$

associated with a potential function G satisfying (4.32) It is easily checked that

$$|M_G(f)(x) - M_G(f)(y)| \quad = \quad |\Psi_G(\delta_x M)(f) - \Psi_G(\delta_y M)(f)| \leq g\ \|\delta_x M - \delta_y M\|_{\mathrm{tv}}$$

from which we conclude that

$$\beta\left(M_G\right) \leq g\ \beta\left(M\right) \tag{4.38}$$

This estimate is clearly only useful when $\beta\left(M\right) < 1/g$.

4.4.3 Normalized Feynman-Kac semigroups

We return to the Feynman-Kac semigroups discussed in section 4.2.1. We let $P_{p,n}$ be the normalized semigroups defined by

$$P_{p,n}(f) := Q_{p,n}(f)/Q_{p,n}(1)$$

For any $0 \leq p \leq q \leq n$, we have

$$P_{p,n}(f) \quad := \quad \frac{Q_{p,q}(Q_{q,n}(f))}{Q_{p,q}(Q_{q,n}(1))} = \frac{Q_{p,q}\left(Q_{q,n}(1)\frac{Q_{q,n}(f)}{Q_{q,n}(1)}\right)}{Q_{p,q}(Q_{q,n}(1))}$$

If we set

$$G_{p,n} = Q_{p,n}(1) \quad \text{and} \quad R^{(n)}_{p,q}(f) := \frac{Q_{p,q}\left(G_{q,n}\ f\right)}{Q_{p,q}(G_{q,n})} = \frac{P_{p,q}\left(G_{q,n}\ f\right)}{P_{p,q}(G_{q,n})}$$

then we find that

$$P_{p,n}(f) \quad = \quad \frac{Q_{p,q}\left(G_{q,n}\ P_{q,n}(f)\right))}{Q_{p,q}(G_{q,n})} = \frac{P_{p,q}\left(G_{q,n}\ P_{q,n}(f)\right))}{P_{p,q}(G_{q,n})} = R^{(n)}_{p,q} P_{q,n}(f)$$

In a more synthetic form, we have proved that

$$\forall 0 \leq p \leq q_1 \leq \ldots \leq q_k \leq n \qquad P_{p,n} = R^{(n)}_{p,q_1} P_{q_1,n} = R^{(n)}_{p,q_1} R^{(n)}_{q_1,q_2} \ldots R^{(n)}_{q_k,n}$$

For instance, for $q = p + 1$ we find that

$$R^{(n)}_{p,p+1}(f) := \frac{Q_{p+1}\left(G_{p+1,n}\ f\right)}{Q_{p+1}(G_{p+1,n})} = \frac{M_{p+1}\left(G_{p+1,n}\ f\right))}{M_{p+1}(G_{p+1,n})}$$

Assuming that

$$M_{p+1}(x, dz) \geq \epsilon\ M_{p+1}(y, dz)$$

for any non negative function f, we find that

$$R^{(n)}_{p,p+1}(f)(x) = \frac{M_{p+1}\left(G_{p+1,n}\ f\right)(x)}{M_{p+1}(G_{p+1,n})(x)} \geq \epsilon^2\ \frac{M_{p+1}\left(G_{p+1,n}\ f\right)(y)}{M_{p+1}(G_{p+1,n})(y)} = R^{(n)}_{p,p+1}(f)(y)$$

This implies that

$$R^{(n)}_{p,p+1}(x,dz) \geq \epsilon^2\ R^{(n)}_{p,p+1}(y,dz) \quad \Longrightarrow \quad \beta\left(R^{(n)}_{p,p+1}\right) \leq 1-\epsilon^2$$

from which we conclude that

$$\beta\left(P_{p,n}\right) \leq \beta\left(R^{(n)}_{p,p+1}\right)\beta\left(P_{p+1,n}\right) \leq (1-\epsilon^2)\ \beta\left(P_{p+1,n}\right) \leq (1-\epsilon^2)^{n-p}$$

4.4.4 Foster-Lyapunov condition

We further assume that E is a topological vector space equipped with the Borel σ-field $\mathcal{E}$.

• **Dobrushin local contraction property**

For any compact subset $C \subset E$, we have

$$\beta(C;M) := \sup_{(x,y)\in C^2} \|M(x,.) - M(y,.)\|_{tv} < 1 \tag{4.39}$$

• **Foster-Lyapunov condition**

There exists some non negative function W on E with compact subset levels, such that

$$M(W)\ \leq\ \epsilon\ W\ +\ c \tag{4.40}$$

for some $\epsilon \in [0,1[$ and some finite constant $c < \infty$. The function W is called a Lyapunov function.

Replacing W by W/c in (4.40) there is no loss of generality to assume that $c = 1$. In addition, replacing W by $W_\epsilon = 1 + \epsilon\ W \geq 1$ we have

$$M\left(W_\epsilon\right) = \epsilon\ M(W)\ +\ 1 \leq \epsilon\ W_\epsilon\ +\ 1$$

Therefore, there is no loss of generality to replace (4.40) by

$$M(W)\ \leq\ \epsilon\ W\ +\ 1 \quad \text{for some function } W \geq 1. \tag{4.41}$$

When the Dobrushin local contraction and the Foster-Lyapunov condition are satisfied, for any $R \in \mathbb{R}_+$, we set

$$\beta^{(R)}(M) \quad = \quad \beta\left(\{W \leq R\}, M\right) = \sup_{(x,y)\ :\ W(x)\vee W(y)\leq R} \|M(x,.) - M(y,.)\|_{tv} \leq 1$$

We consider a Markov transition M on some complete separable metric space E such that

$$M(x,dy) \geq m(x,y)\ \lambda(dy) \tag{4.42}$$

for some some strictly positive Radon measure $\lambda(dy)$ (i.e. nonempty open balls have positive measure). We further assume that for any compact set A there exists some positive measurable function q_A s.t.

$$\inf_{x\in A} m(x,y) \geq q_A(y) \tag{4.43}$$

In this situation, the Dobrushin local contraction condition (4.39) is satisfied. For instance, the minorization condition (4.43) is met when the functions $m(x,y)$ are lower semicontinuous w.r.t. the first

variable, and upper semicontinuous w.r.t. the second. We recall that a characteristic of Radon measures is that the measure of a Borel set B is the supremum of the measures $\lambda(A)$ of the compact sets $A \subset B$. Thus, if λ is a strictly positive Radon measure on E, one can always find for every open set $B \subset E$ a compact $A \subset B$ such that

$$\lambda(A) \geq (1/2)\ \lambda(B)\ (> 0)$$

This ensures that λ charges all the compact sets.

To check this claim, it is of course tempting to set $\inf_{x\in A} m(x,y) = q_A(y)$ but it is well known that the infimum of an uncountable collection of functions may fails to be measurable. Under the condition (4.43), we clearly have that

$$\forall x \in A \qquad M(x,dy) \geq q_A(y)\ \lambda(dy) := \gamma_A(dy) \quad \text{with} \quad \gamma_A(1) = \lambda(q_A) > 0$$

In this situation, the Dobrushin condition (4.39) is clearly met with

$$\forall x \in A \qquad M(x,dy) \geq \epsilon_A\ \nu_A(dy)$$

with

$$\epsilon_A = \gamma_A(1) > 0 \quad \text{and} \quad \nu_A(dy) := \gamma_A(dy)/\gamma_A(1)$$

When the density function $m(x,y)$ is lower semicontinuous w.r.t. the first variable and upper semicontinuous w.r.t. the second variable, there exists some measurable function $h_A \ : \ y \mapsto h_A(y)$ such that

$$\inf_{x\in A} m(x,y) = m(h_A(y),y) := q_A(y) > 0$$

A proof of this result can be found in [115]. It this situation, the minorization condition (4.43) is clearly met. This ends the proof of the assertion.

The Foster-Lyapunov condition ensures that the Markov chain X_n with transition probabilities M has little chances to escape from the level sets $\{W \leq w\}$ of the function W. Indeed, we have

$$(4.40) \Longrightarrow \ M^n(W) \ \leq \ \epsilon^n\ W + c\ (1 + \epsilon + \ldots + \epsilon^{n-1}) \leq \epsilon^n\ W + c/(1-\epsilon)$$

from which we prove the uniform estimate

$$\sup_{n\geq 0} \mathbb{E}\left(W(X_n)\right) \leq \epsilon^n\ \mathbb{E}(W(X_0)) + c/(1-\epsilon) \leq C := \mathbb{E}(W(X_0)) + c/(1-\epsilon)$$

Using Markov inequality, when the level sets of W are compact we have

$$\forall \rho > 0 \qquad \exists \{W \leq C/\rho\} := A_\rho \text{ compact} \quad \text{s.t.} \quad \sup_{n\geq 0} \mathbb{P}\left(W(X_n) \notin A_\rho\right) \leq \rho$$

We let V be a non negative function V s.t.

$$M(V) \leq c_1\ V + c_2$$

for some $c_1, c_2 \geq 0$. We equip the set of probability measures $\mathcal{P}(S)$ on some state space E with the V-norm.

The V-Dobrushin ergodic coefficient $\beta_V(M)$ is defined by

$$\beta_V(M) \ = \ \sup\left\{\mathrm{osc}_V(M(f))\ ,\ f \ : \ \mathrm{osc}_V(f) \leq 1\right\} = \sup_{(x,y)\in E^2} \frac{\|M(x,.) - M(y,.)\|_V}{1 + [V(x) + V(y)]}$$

with $\|.\|_V$ and $\mathrm{osc}_V(.)$ defined in (12) and (13).

The r.h.s. formulation in the above display is readily checked using the fact that

$$\beta_V(M) \quad = \quad \sup_{(x,y)\in E^2} \quad \sup_{f:\mathrm{osc}_V(f)\leq 1} \frac{|M(f)(x) - M(f)(y)|}{1 + [V(x) + V(y)]}$$

Arguing as in the proof of (4.28), for any couple of measures $\mu_1, \mu_2 \in \mathcal{P}(S)$, and any function f s.t. $\|f\|_V < \infty$, and any $n \in \mathbb{N}$ we have

$$\begin{aligned} \mathrm{osc}_V(M(f)) &\leq \beta_V(M)\ \mathrm{osc}_V(f) \\ \|\mu_1 M - \mu_2 M\|_V &\leq \beta_V(M)\ \|\mu_1 - \mu_2\|_V \quad \text{and} \quad \beta_V(M^n) \leq \beta_V(M)^n \end{aligned} \tag{4.44}$$

When the drift condition (4.40) and the local Dobrushin condition (4.39) are satisfied for some function W and some parameter $\epsilon \in [0, 1[$, there exist some positive function V s.t. $\beta_V(M) < 1$.

4.5 Some illustrations

4.5.1 Minorization condition

Let a be a bounded function on $E = \mathbb{R}$, and let M be the Markov transition associated with the evolution equation

$$X_n = a(X_{n-1}) + W_n$$

where W_n stands for a sequence of independent and absolutely continuous r.v. with common density $p(w) = \frac{1}{2\lambda}\ e^{-\lambda|w|}$. In this situation, if we fix a point $x_0 \in E$, we have

$$\nu(dy) := M(x_0, dy) \leq M(x, dy)\ e^{\lambda||y-a(x)|-|y-a(x_0)||} \leq M(x, dy)\ e^{\lambda\ \mathrm{osc}(a)}$$

This implies that

$$M(x, dy)\ \geq\ \epsilon\ \nu(dy) \quad \text{with} \quad \epsilon = \exp\left(-\lambda_{\mathrm{osc}}(a)\right)$$

We consider a compact set $E' \subset E = \mathbb{R}^d$, and we let $p(x, y)$ some continuous positive function on $(\mathbb{R}^d \times \mathbb{R}^d)$. The Markov transition $M(x, dy) \propto p(x, y)\ 1_{S'}(y)\ dy$ on E' satisfies (4.29). We check this claim using the fact that

$$\sqrt{\epsilon} \leq \frac{p(x', y)}{p(x, y)} \leq \frac{1}{\sqrt{\epsilon}} \quad \text{with} \quad \sqrt{\epsilon} := \inf_{x,x',y\in S'} \frac{p(x', y)}{p(x, y)} > 0 \tag{4.45}$$

These estimates implies (4.29) with $\nu(dx) \propto M(x_0, dy)$. This indicates that (4.29) is satisfied for any Markov transitions with a continuous density on some compact space (equipped with some metric). For instance any regular Markov chain evolving in our galaxy satisfies (4.29).

We assume that there exist some subset $A \subset E$ and some positive measure γ s.t. $\gamma(A) > 0$ and for any $x \in E$

$$M(x, dy)\ 1_A(y) \geq \gamma(dy)\ 1_A(y) \tag{4.46}$$

In this case, we have

$$M(x, dy) \geq\ M(x, dy)\ 1_A(y) \geq \gamma(dy)\ 1_A(y) = \epsilon\ \nu(dy)$$

with

$$\epsilon = \gamma(A) \quad \text{and} \quad \nu(dy) = \frac{\gamma(dy)1_A(y)}{\gamma(A)}$$

For instance, the Gaussian transition on $E = \mathbb{R}$ defined by

$$M(x, dy) = \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp\left(-\frac{1}{2\sigma^2(x)}(y - a(x))^2\right) dy$$

satisfies (4.46) with $A = \mathbb{R}$ as soon as

$$0 < \sigma^2_{min} \leq \sigma^2(x) \leq \sigma^2_{max} < \infty \quad \text{and} \quad \|a\| < \infty$$

We check this claim using the fact that

$$\begin{aligned} y \geq 0 \quad &\Rightarrow \quad \sup_{x\in S}(y - a(x))^2 \leq y^2 + 2y\|a\|_A + \|a\|^2 = (y + \|a\|)^2 \\ y \leq 0 \quad &\Rightarrow \quad \sup_{x\in S}(y - a(x))^2 \leq y^2 - 2y\|a\|_A + \|a\|^2 = (y - \|a\|)^2 \end{aligned}$$

and

$$M(x, dy) \geq \gamma(dy) := \gamma_1(dy) + \gamma_2(dy)$$

with

$$\begin{aligned} \gamma_1(dy) &= \frac{1}{\sqrt{2\pi\sigma^2_{max}}} \exp\left(-\frac{1}{2\sigma^2_{min}}(y + \|a\|)^2\right) 1_{y\geq 0}\, dy \\ \gamma_2(dy) &= \frac{1}{\sqrt{2\pi\sigma^2_{max}}} \exp\left(-\frac{1}{2\sigma^2_{min}}(y - \|a\|)^2\right) 1_{y<0}\, dy \end{aligned}$$

Notice that in this case we have

$$\begin{aligned} \gamma_1(1) + \gamma_2(1) &= \frac{\sigma_{min}}{\sigma_{max}} \left[\mathbb{P}\left(-\|a\| + \sigma_{min}Y \geq 0\right) + \mathbb{P}\left(\|a\| + \sigma_{min}Y \leq 0\right)\right] \\ &= \frac{\sigma_{min}}{\sigma_{max}} \left[1 - \mathbb{P}\left(|Y| \leq \|a\|/\sigma_{min}\right)\right] > 0 \end{aligned}$$

where Y stands for a centered Gaussian r.v. with unit variance.

4.5.2 Gaussian transitions

A simple way to check the Dobrushin local contraction condition is to prove that for any compact subset $C \subset E$, there exists some $\epsilon_C \in]0, 1]$ and some probability measure ν_C on E such that

$$\forall x \in C \quad M(x, dy) \geq \epsilon_C\, \nu_C(dy) \tag{4.47}$$

In this situation, using the same arguments as the ones we used in (4.29) we prove that

$$\beta(C; M) \leq (1 - \epsilon_C)$$

In the literature on Markov chain stability the subsets C satisfying the minorization condition (4.47) are often called "small" sets.

This local contraction condition is satisfied for most of the Markov chain encountered in practice. For instance, for the Gaussian transition

$$M(x, dy) = \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp\left(-\frac{1}{2\sigma^2(x)}(y - a(x))^2\right) dy$$

associated with some *locally bounded* drift and variance functions a and σ^2 on $E = \mathbb{R}$ we have

$$\begin{aligned} y \geq 0 \quad &\Rightarrow \quad \sup_{x\in A}(y - a(x))^2 \leq y^2 + 2y\|a\|_A + \|a\|^2_A = (y + \|a\|_A)^2 \\ y \leq 0 \quad &\Rightarrow \quad \sup_{x\in A}(y - a(x))^2 \leq y^2 - 2y\|a\|_A + \|a\|^2_A = (y - \|a\|_A)^2 \end{aligned}$$

for any bounded subset $A \subset E$, with $\|a\|_A := \sup_{x\in A} |a(x)|$. We further assume that

$$\forall x \in A \qquad 0 < \sigma^2_{min,A} \leq \sigma^2(x) \leq \sigma^2_{max,A} < \infty$$

This implies that

$$\begin{aligned} M(x,dy) &\geq \gamma(dy) \\ &:= 1_{y\geq 0} \left(\sqrt{2\pi\sigma^2_{max,A}}\right)^{-1} \exp\left(-\frac{1}{2\sigma^2_{min,A}}(y+\|a\|_A)^2\right) dy \\ &\quad +1_{y<0} \left(\sqrt{2\pi\sigma^2_{max,A}}\right)^{-1} \exp\left(-\frac{1}{2\sigma^2_{min,A}}(y-\|a\|_A)^2\right) dy \;\geq\; \tau_A\, \nu(dy) \end{aligned}$$

with the probability measure $\nu(dy) = \gamma(dy)/\gamma(1)$ and the $]0,1[$ valued constant

$$\tau_A = \frac{\sigma_{min,A}}{\sigma_{max,A}} \left[1 - \int_{-\|a\|_A/\sigma_{min,A}}^{\|a\|_A/\sigma_{min,A}} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy\right]$$

This clearly implies that

$$\sup_{(x,y)\in A^2} \|M(x,.) - M(y,.)\|_{tv} \leq 1 - \tau_A$$

In the Gaussian model discussed above, we notice that

$$|a(x)| \leq \epsilon\, |x| \Rightarrow \mathbb{E}\left[\; |X_n| \;\; |X_{n-1} = x\right] \leq \epsilon\, |x| + \sigma^2$$

In this case, $W(x) = |x|$ satisfies (4.40) with $c = \sigma^2$.

More generally, suppose that X_n and X'_n are two non necessarily independent copies of the transition of the chain starting at x and x_0; that is, we have that

$$\begin{aligned} M(x,dy) &:= \mathbb{P}\left(X_n \in dy \mid X_{n-1} = x\right) \\ M(x_0,dy) &:= \mathbb{P}\left(X_n \in dy \mid X_{n-1} = x_0\right) \end{aligned}$$

We suppose that the state space E is equipped with some metric d and we have the local contraction inequality

$$\mathbb{E}\left(d(X_n, X'_n) \mid (X_{n-1}, X'_{n-1}) = (x, x_0)\right) \leq \epsilon\, d(x, x_0)$$

Returning to the Gaussian model discussed above, we can take

$$X_n = a(x) + \sigma\, Y \quad \text{and} \quad X'_n = a(x_0) + \sigma\, Y$$

where Y stands for a centered Gaussian r.v. with unit variance. In this situation, the local contraction stated above is met for the Euclidian distance $d(x,y) = |x-y|$ as soon as

$$|a(x) - a(x_0)| \leq \epsilon\, |x - x_0|$$

We set $W(x) := d(x, x_0)$ for some fixed state $x_0 \in E$. Using the triangle inequality

$$d(X_n, x_0) - d(X'_n, x_0) \leq d(X_n, X'_n)$$

we prove that

$$\mathbb{E}\left(d(X_n, x_0) - d(X'_n, x_0) \mid (X_{n-1}, X'_{n-1}) = (x, x_0)\right)$$

$$= M(W)(x) - M(W)(x_0) \leq \epsilon\, d(x, x_0) = \epsilon\, W(x)$$

This implies that the Foster-Lyapunov condition is met with

$$M(W)(x) \leq \epsilon\, W(x) + c \quad \text{with} \quad c := M(W)(x_0)$$

We end this section with a sufficient condition of the Foster-Lyapunov condition. Suppose there exists some subset $A \subset E$ s.t.

$$\left\{ \begin{array}{lcll} M(W)(x) & \leq & \epsilon\, W(x) & \text{for any} \quad x \in E - A \\ M(W)(x) & \leq & c & \text{for any} \quad x \in A \end{array} \right. \tag{4.48}$$

then we have

$$\forall x \in S \quad M(W)(x) \leq \epsilon\, W(x)\, 1_{E-A}(x) \; + \; c\, 1_A(x) \leq \epsilon W \; + \; c$$

Whenever $M(W)$ is continuous the condition (4.48) is met as soon as we can find some compact set A s.t.

$$\forall x \notin A \qquad M(W)(x) \leq \; \epsilon\, W(x)$$

for some $\epsilon \in [0, 1[$. In this case (4.48) is satisfied with $c = \sup_{x \in A} |W(x)|$.

4.5.3 Some quantitative estimates

In the further development of this section, we assume that the condition (4.41) is met and we set $V_\rho = \rho\, W$, for some $\rho \in]0, 1]$. Notice that

$$(4.41) \;\Rightarrow\; M(V_\rho) \;\leq\; \epsilon\, V_\rho \;+\; \rho$$

In addition, we have the uniform estimate

$$(V_\rho = \rho\, W \quad \text{and} \quad W \geq 1) \;\Rightarrow\; V_\rho^{-1} M(V_\rho) \;\leq\; \epsilon \;+\; \frac{\rho}{\rho W} \leq 1 + \epsilon$$

We also notice that for any $R \geq 1$ we have

$$W(x) \geq R \;\Rightarrow\; \left\{ \begin{array}{rcl} V_\rho(x)^{-1}\, M(V_\rho)(x) & \leq & \epsilon \;+\; \frac{1}{W(x)} \leq \epsilon + \frac{1}{R} \\ V_\rho(x) & = & \rho\, W(x) \geq \rho R \end{array} \right. \tag{4.49}$$

and

$$W(x) \leq R \;\Rightarrow\; \left\{ \begin{array}{rcl} V_\rho(x)^{-1}\, M(V_\rho)(x) & \leq & 1 + \epsilon \\ \rho \leq V_\rho(x) & = & \rho\, W(x) \leq \rho R \end{array} \right. \tag{4.50}$$

We set

$$\Delta_\rho(x, y) := \frac{\|M(x, .) - M(y, .)\|_{V_\rho}}{1 + V_\rho(x) + V_\rho(y)}$$

By definition of the V-norm (12), using the triangle inequality (for the the total variation of the measures; namely $|\mu_1 + \mu_2| \leq |\mu_1| + |\mu_2|$, for any measures μ_1, μ_2), we prove the following decomposition

$$\begin{aligned} &\Delta_\rho(x, y) \\ &= \frac{1}{1 + V_\rho(x) + V_\rho(y)}\; \|M(x, .) - M(y, .)\|_{tv} \\ &\qquad + \frac{V_\rho(x)}{1 + V_\rho(x) + V_\rho(y)}\; \frac{M(V_\rho)(x)}{V_\rho(x)} + \frac{V_\rho(y)}{1 + V_\rho(x) + V_\rho(y)}\; \frac{M(V_\rho)(y)}{V_\rho(y)} \\ &\leq \frac{1}{1 + V_\rho(x) + V_\rho(y)}\; \|M(x, .) - M(y, .)\|_{tv} \\ &\qquad + \frac{V_\rho(x) + V_\rho(y)}{1 + V_\rho(x) + V_\rho(y)}\; \left(\frac{M(V_\rho)(x)}{V_\rho(x)} \vee \frac{M(V_\rho)(y)}{V_\rho(y)} \right) \end{aligned}$$

When $W(x) \wedge W(y) \geq R$, using (4.49) we find that

$$\begin{aligned}\Delta_\rho(x,y) &\leq \frac{1}{1+V_\rho(x)+V_\rho(y)} + \frac{V_\rho(x)+V_\rho(y)}{1+V_\rho(x)+V_\rho(y)} \left(\epsilon + \frac{1}{R}\right) \\ &= 1 - \left(1 - \frac{1}{1+V_\rho(x)+V_\rho(y)}\right) \left(1 - \left(\epsilon + \frac{1}{R}\right)\right)\end{aligned}$$

from which we conclude that

$$\sup_{W(x)\wedge W(y)\geq R} \Delta_\rho(x,y) \leq 1 - \left(1 - \frac{1}{1+2\rho R}\right) \left(1 - \left(\epsilon + \frac{1}{R}\right)\right) < 1$$

for any $\rho \in]0,1]$. Using (4.50) we also find that

$$\begin{aligned}\sup_{W(x)\vee W(y)\leq R} \Delta_\rho(x,y) &\leq \frac{1}{1+2\rho}\, \beta^{(R)}(M) + 2\, \frac{\rho R}{1+2\rho}\, (1+\epsilon) \\ &\leq \beta^{(R)}(M) + 4\rho R < 1\end{aligned}$$

for any $\rho < (1-\beta^{(R)}(M))/(4R)$.

Combining these estimates with (4.44) we readily prove the following result.

When the drift condition (4.40) and the local Dobrushin condition (4.39) are satisfied for some function W and some parameter $\epsilon \in [0,1[$, there exist some positive function V s.t.

$$\beta_V(M) < 1$$

In this case there exists an unique invariant measure $\pi = \pi M$ and we have the exponential contraction inequality

$$\|\mu_1 M^n - \mu_2 M^n\|_V \leq \beta_V(M)^n\, \|\mu_1 - \mu_2\|_V \longrightarrow_{n\uparrow\infty} 0 \tag{4.51}$$

In exercise 4.7.1 we design a function V such that

$$\beta_V(M) \leq 1 - \frac{1-\beta^{(R_\epsilon)}(M)}{R_\epsilon(1+2\sqrt{3})} \quad \text{with} \quad R_\epsilon := 2/(1-\epsilon)$$

By definition of $\beta_V(M^n)$, for any function f such that

$$|f(x)| \leq 1/2 + V(x) \quad (\Rightarrow\ |f(x)-f(y)| \leq 1 + V(x) + V(y))$$

and for any $(x,y) \in E$ we have

$$|M^n(f)(x) - M^n(f)(y)| \leq \beta_V(M^n)\ (1 + V(x) + V(y))$$

This implies that

$$\begin{aligned}|M^n(f)(x) - \pi(f)| &\leq \int \pi(dy)\ |M^n(f)(x) - M^n(f)(y)| \\ &\leq \beta_V(M^n)\ (1 + V(x) + \pi(V))\end{aligned}$$

4.6 The ergodic theorem

We consider the occupation measures of a Markov chain X_n with transition K on some state space E defined by

$$\pi^n := \frac{1}{n+1} \sum_{0 \leq p \leq n} \delta_{X_p} = \pi + \frac{1}{\sqrt{n+1}}\, V^n \ \left(\Leftrightarrow \ V^n = \sqrt{n+1}\, [\pi^n - \pi] \right) \tag{4.52}$$

We further assume that the Dobrushin ergodic coefficient of K is s.t.

$$\beta(K^n) \leq a\ e^{-b\ n} \tag{4.53}$$

for some parameters $a < \infty$, and $b > 0$, and for any $n \geq 0$. In this situation, the chain an unique invariant measure $\pi = \pi K$ and we have $\text{osc}\,(K^n(f)) \leq a\ e^{-b\ n}$, for any f s.t. $\text{osc}(f) \leq 1$. Using the fact that

$$\|K^n(f) - \pi(f)\| = \|K^n(f) - \pi K^n(f)\| \leq \text{osc}\,(K^n(f))$$

we check that the functional series

$$P(f)(x) = \sum_{n \geq 0} K^n(f)(x) \iff P := [Id - K]^{-1} = \sum_{n \geq 0} K^n$$

are well defined bounded functions for any f s.t. $\pi(f) = 0$, and $\text{osc}\,(f) \leq 1$. We check this claim using the fact that

$$\|P(f)\| \leq \sum_{n \geq 0} \|K^n(f) - \pi K^n(f)\| \leq \sum_{n \geq 0} \text{osc}\,(K^n(f)) \leq a/1 - e^{-b}$$

In addition, they solve the Poisson equation

$$g = P(f) \ \Rightarrow \ -L(g) = f \quad \text{with} \quad L = K - Id \tag{4.54}$$

for any given function f s.t. $\pi(f) = 0$. We check this claim using the fact that

$$[Id - K]P(f)(x) = \sum_{n \geq 0} K^n(f)(x) - \sum_{n \geq 1} K^n(f)(x) = f(x)$$

By construction, we have $[Id - K](g) = f - \pi(f)$, from which we check that

$$\begin{aligned} f(X_p) - \pi(f) &= [Id - K](g)(X_p) \\ &= g(X_p) - K(g)(X_p) = g(X_p) - \mathbb{E}\,(g(X_{p+1}) \mid X_p) \\ &= -\underbrace{(g(X_{p+1}) - g(X_p))}_{\Delta g(X_{p+1})} + \underbrace{(g(X_{p+1}) - \mathbb{E}\,(g(X_{p+1}) \mid X_p))}_{= \Delta \mathcal{M}_{p+1}(g)} \end{aligned}$$

This yields the decomposition

$$V^n(f) = \frac{1}{\sqrt{n+1}}\, (g(X_{n+1}) - g(X_0)) - \frac{1}{\sqrt{n+1}}\, \mathcal{M}_{n+1}(g) = -\frac{\mathcal{M}_{n+1}(g)}{\sqrt{n+1}} + \text{O}(1/\sqrt{n}) \tag{4.55}$$

A a direct consequence of (4.55), we have $|\mathbb{E}[V^n(f)]| \to_{n\uparrow\infty} 0$, and

$$\mathbb{E}\left(V^n(f)^2\right) = \frac{1}{n+1} \underbrace{\mathbb{E}\left(\langle \mathcal{M}(g)\rangle_{n+1}\right)}_{\leq (n+1)\ \mathrm{osc(g)}^2} + \mathrm{o}(1/n) \Rightarrow \sup_{n\geq 0} \mathbb{E}\left(V^n(f)^2\right) < \infty \tag{4.56}$$

To compute the limiting variance, we notice that

$$\begin{aligned} n^{-1}\,\mathbb{E}\left(\langle \mathcal{M}(g)\rangle_n\right) &= \mathbb{E}\left(n^{-1}\sum_{0<p\leq n}\left[K(g^2)-K(g)^2\right](X_{n-1})\right) \\ &= \mathbb{E}\left(\pi^{n-1}(K(g^2)-K(g)^2)\right) \longrightarrow_{n\to\infty} \pi\left(K(g^2)-K(g)^2\right) \end{aligned}$$

Thus, recalling that $\pi K = \pi$, we also have the asymptotic result

$$\mathbb{E}\left(\left[V^n(f)^2\right]\right) \longrightarrow_{n\to\infty} \sigma^2(f) := \pi\left(g^2\right) - \pi\left(K(g)^2\right)$$

Since g satisfies the Poisson equation we have

$$g = K(g) + (f - \pi(f))$$

$$\Rightarrow \pi(g^2) = \pi(K(g)^2) + 2\,\pi((f-\pi(f))K(g)) + \pi((f-\pi(f))^2)$$

This implies that

$$\begin{aligned} \sigma^2(f) &= 2\pi((f-\pi(f))K(g)) + \pi((f-\pi(f))^2) \\ &= 2\sum_{p\geq 1}\pi([f-\pi(f)]K^p[f-\pi(f)]) + \pi([f-\pi(f)]^2) \end{aligned}$$

4.7 Exercises

Exercise 4.7.1 *The aim of this exercise is to quantify more explicitly the geometric drift contraction inequalities discussed in section 4.4.4. We set $R_\epsilon := 2/(1-\epsilon)$, $\alpha_\epsilon := 1 - \beta^{(R_\epsilon)}(M)$, and $\delta := (1-\epsilon)$.*

- *Choosing $R = R_\epsilon$ in (4.49) and (4.50), check that*

$$\forall \rho \in]0,1] \qquad \sup_{W(x)\wedge W(y)\geq R_\epsilon} \Delta_\rho(x,y) \leq 1 - \frac{1}{2}\,\frac{4\rho(1-\epsilon)}{(1-\epsilon)+4\rho} < 1$$

and

$$\forall \rho \in]0, \alpha_\epsilon\delta/8[\qquad \sup_{W(x)\vee W(y)\leq R_\epsilon} \Delta_\rho(x,y) \leq 1 - \left(\alpha_\epsilon - \frac{8\rho}{1-\epsilon}\right) < 1$$

- *We set $u := 4\rho/\delta$, and*

$$\begin{aligned} g(u) &:= \frac{1}{2}\,\frac{4\rho(1-\epsilon)}{(1-\epsilon)+4\rho} = \frac{\delta}{2}\left(1 - \frac{1}{1+u}\right) \\ h(u) &:= \left(\alpha_\epsilon - \frac{8\rho}{1-\epsilon}\right) = (\alpha_\epsilon - 2u) \end{aligned}$$

Check that these two functions intersects at some point

$$u = \sqrt{a^2+b} - a \in [0,b]$$

with

$$a := \frac{1}{2}\left(1 - b + \frac{\delta}{4}\right) \leq \frac{1}{2} \quad \text{and} \quad b := \frac{\alpha_\epsilon}{2}$$

Prove that for any $v \geq 0$ *we have*

$$\sqrt{1+v} \geq 1 + \frac{v}{2\sqrt{1+v}}$$

and deduce that

$$g(u) = h(u) \geq \frac{\delta b}{1 + 2\sqrt{3}}$$

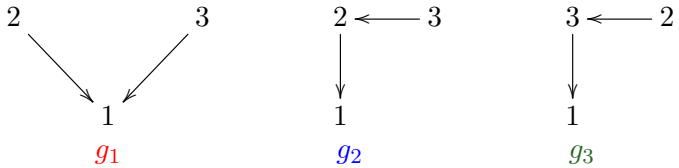
- *Deduce from the above that for* $\rho = u\delta/4$ *we have*

$$\beta_{V_\rho}(M) \leq 1 - \frac{(1-\epsilon)(1-\beta^{(R_\epsilon)}(M))}{2(1+2\sqrt{3})}$$

Exercise 4.7.2 *We consider a stochastic matrix with positive entries*

$$M = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

For any $x \in S = \{1, 2, 3\}$*, we denote by* $\mathcal{G}(i)$ *the set of all the* i*-graphs. An* i*-graph is a set of directed edges without any loops connecting all the states* $j \neq i$ *without cycles to* i*; and with a single edge starting from the states* $j \neq i$*. For instance, the set* 1*-graphs* $\mathcal{G}(1) = \{g_1, g_2, g_3\}$ *with the directed graphs defined below*



g_1 g_2 g_3 *and*

Prove that the unique invariant measure of the chain is given by

$$\pi(i) = \gamma(i) / \sum_{1 \leq j \leq 3} \gamma(j) \quad \text{with} \quad \gamma(i) = \sum_{g \in \mathcal{G}(i)} \prod_{(k,l) \in g} p_{k,l}$$

For instance, for $i = 1$ *we have*

$$\gamma(1) = \sum_{g \in \mathcal{G}(1)} \prod_{(i,j) \in g} p_{i,j} = (p_{21}p_{31} + p_{32}p_{21} + p_{23}p_{31})$$

Exercise 4.7.3 *We consider a Markov transition on some finite state space such that the* $M^m(x, y) > 0$*, for any* $x, y \in E$*, for some* $m \geq 1$*. Prove that the unique invariant measure of the chain is given by*

$$\pi(x) = \gamma(x) / \sum_{y \in E} \gamma(y) \quad \text{with} \quad \gamma(x) = \sum_{g \in \mathcal{G}(x)} \prod_{(y,z) \in g} M(y, z)$$

where $\mathcal{G}(x)$ *stands for the set of* x*-graphs defined in exercise 4.7.2.*

Exercise 4.7.4 *We consider the stochastic matrix presented in exercise 4.7.2.*

- *Prove that the characteristic polynomial* $P(\lambda) = \mathrm{Det}(M - \lambda I)$ *is given by*

$$P(\lambda) = (1-\lambda) \ \left(\lambda^2 + (1-A)\lambda + C\right)$$

with $C = 1 - (A + B)$, *and*

$$\begin{aligned} A &= p_{11} + p_{22} + p_{33} \\ B &= p_{23}p_{32} + p_{12}p_{21} + p_{13}p_{31} - (p_{11}p_{22} + p_{11}p_{33} + p_{22}p_{33}) \end{aligned}$$

- *Check that*

$$-\frac{1-A}{2} = 1 - (q_{12} + q_{13} + q_{23})$$

with the parameters $q_{i,j} = (p_{ij} + p_{ji})/2$.

- *Check that*

$$B = -3 + 4 \ (q_{12} + q_{13} + q_{23}) - D$$

with

$$D = 4\,(q_{12}q_{13} + q_{12}q_{23} + q_{13}q_{23}) - (p_{21}p_{23} + p_{12}p_{13} + p_{31}p_{32})$$

- *Deduce that*

$$\left(\frac{1-A}{2}\right)^2 - C = \Delta(q) + \delta(p)$$

with the parameters

$$\begin{aligned} \Delta(q) &= \frac{1}{2}\left[(q_{12} - q_{13})^2 + (q_{12} - q_{23})^2 + (q_{13} - q_{23})^2\right] \\ \delta(p) &= [p_{12}p_{13} - q_{12}q_{13}] + [p_{21}p_{23} - q_{21}q_{23}] + [p_{31}p_{32} - q_{31}q_{32}] \end{aligned}$$

- *Conclude that the eigenvalues of* M *are given by* $\lambda_1 = 1$, *and*

$$\begin{aligned} \lambda_2 &= (1 - (q_{12} + q_{13} + q_{23})) + \sqrt{\Delta(q) + \delta(p)} \\ \lambda_3 &= (1 - (q_{12} + q_{13} + q_{23})) - \sqrt{\Delta(q) + \delta(p)} \end{aligned}$$

with the convention $\sqrt{-a} = i\sqrt{a}$, *for any* $a \geq 0$.

- *In the reversible case, check that* $\delta(p) = 0$, *and*

$$\lambda_3 \leq \lambda_2 \leq \lambda_1 = 1$$

Chapter 5

Continuous time models

5.1 Continuous vs discrete time models

The abstract Markov chain framework introduced in Section 4 encapsulates continuous time evolutions of nonanticipative processes, as well as càdlàg Markov processes $(X'_t)_{t\geq 0}$, taking values on some Polish state space E (i.e., a separable and completely metrizable topological space).

For instance, given some time mesh $(t_n)_{n\geq 0}$, the sequence of random variables X_n defined by

$$X_n = X'_{t_n} \ , \qquad X_n = (X'_t)_{t_n\leq t\leq t_{n+1}} \ , \quad \text{and resp.} \quad X_n = (X'_t)_{0\leq t\leq t_n}$$

are Markov chains taking values in $E_n = E$, $E_n = D([t_n, t_{n+1}], E)$, and resp. $E_n = D([0, t_n], E)$, where $D([a, b], E)$ stands for the set of càdlàg paths from the time interval $[a, b]$ into the set E.

Of course the Monte Carlo simulation of continuous time models $(X'_t)_{t\geq 0}$ often requires to discretize the time into small time intervals, just like a video is actually discretized into a series of discrete generation frame sequences.

For instance, let us suppose we are given an $\mathbb{R}^d$-valued Itô stochastic differential equation

$$dX'_t = a'_t(X'_t)\ dt + \sigma'_t(X'_t)\ dW_t \tag{5.1}$$

with some initial random variable X'_0 with distribution $\eta'_0 = \mathrm{Law}(X'_0)$. In the above display, W_t is a standard d-dimensional Wiener process, and for any $x \in \mathbb{R}^d$, $\sigma'_t(x) = (\sigma'_{t,i,j}(x))_{1\leq i,j\leq d}$ and $a'_t(x) = (a'_{t,i}(x))_{1\leq i\leq d}$ are, respectively, a symmetric nonnegative definite matrix and a vector with appropriate dimensions. In the further development of these lectures notes, it is implicitly assumed that the drift and diffusion functions are sufficiently regular so that the stochastic processes are well defined at all times.

A Euler type discretization $X_n \simeq X'_{t_n}$ of the diffusion (5.1) is given by the Markov chain

$$X_n = X_{n-1} + a_n\left(X_{n-1}\right)\ \Delta + \sigma_n\left(X_{n-1}\right)\ \left(W_{t_n} - W_{t_{n-1}}\right) \tag{5.2}$$

on the time mesh $(t_n)_{n\geq 0}$ s.t. $(t_n - t_{n-1}) = \Delta > 0$, with the drift and diffusion functions given by

$$(a_n, \sigma_n) = \left(a'_{t_{n-1}}, \sigma'_{t_{n-1}}\right)$$

and the initial random variable, $X_0 = X'_0$.

High order or implicit time discretization schemes can also be used to cure the instability of some evolution models.

In the same vein, suppose we are given a jump type Markov process X'_t that evolves between jump times T_n as in (5.1).

The jump times T_n are defined in terms of a sequence $(e_n)_{n\geq 1}$ of independent and identically distributed *(abbreviated i.i.d.)* exponentially distributed random variables with unit parameter. The successive jump times are defined sequentially by the following recursion

$$T_n = \inf\left\{ t \geq T_{n-1} \ : \ \int_{T_{n-1}}^{t} \lambda'_u(X'_u)\ du \geq e_n \right\} \tag{5.3}$$

with $T_0 = 0$, and some nonnegative function λ'_u. At jump times T_n, the process at X'_{T_n-} jumps to a new location X'_{T_n} randomly selected with some distribution $S'_{T_n-}(X'_{T_n-}, dy)$.

In the above discussion, we have implicitly assumed that the parameters $(a'_t, \sigma'_t, \lambda'_t)$ are sufficiently regular, so that the jump diffusion defined by (5.1) and (5.3) is well defined on the real line $\mathbb{R}_+ = [0, \infty[$.

Discretizing the integral in (5.3) on a time mesh t_n with time step h, we define a sequence of random times

$$\begin{aligned} T_n^h &= \inf\left\{ t_m \geq T_{n-1}^h \ : \ \exp\left(- \sum_{T_{n-1}^h \leq t_k < t_m} \lambda'_{t_k}(X'_{t_k})\ h \right) \leq -\exp e_n = u_n \right\} \\ &= \inf\left\{ t_m \geq T_{n-1}^h \ : \ \prod_{T_{n-1}^h \leq t_k < t_m} e^{-\lambda'_{t_k}(X'_{t_k})\ h} \ \leq u_n \right\} \end{aligned}$$

where u_n stands for a sequence of i.i.d. random variables on $[0, 1]$.

This shows that

$$\mathbb{P}\left(T_n^h = t_m \mid X'_{t_k},\ t_k < t_m \right) = \left(\prod_{T_{n-1}^h \leq t_k < t_{m-1}} e^{-\lambda'_{t_k}(X'_{t_k})\ h} \right) \left(1 - e^{-\lambda'_{t_{m-1}}(X'_{t_{m-1}})\ h} \right)$$

Thus, a discrete time approximation of the jump-diffusion model on a time mesh t_n is given by a Markov chain X_n with elementary transitions

$$K_n(x, dz) = \int M_n(x, dy)\ J_n(y, dz) \tag{5.4}$$

where M_n stands for the transition of the Markov chain (5.2), and J_n is the geometric jump type Markov transition defined in terms of the couple $(G_n, S_n) = \left(\exp\left(-\Delta\lambda'_{t_n}\right), S'_{t_n}\right)$, by the following equation

$$J_n(y, dz) = G_n(y)\ \delta_y(dz) + (1 - G_n(y))\ \ S_n(y, dz) \tag{5.5}$$

We refer to the articles by E. Pardoux and D. Talay [486], V. Bally and D. Talay [34], as well as the seminal articles by P.E. Kloeden, E. Platen, and their co-authors [384, 385, 386, 387], for a detailed convergence analysis of time discretization schemes.

5.2 Martingale decomposition

To clarify the presentation, we suppress the index $(.)'$ and we let X_t be a jump-diffusion process defined as in (5.1) and (5.3) replacing $(a'_t, \sigma'_t, \lambda'_t, S'_t)$ by some parameters $(a_t, \sigma_t, \lambda_t, S_t)$. For any $s \in \mathbb{R}_+$ and $x \in \mathbb{R}$ we denote by

$$t \in [s, \infty[\ \mapsto \ \varphi_{s,t}(x) \in \mathbb{R}^d$$

the solution of the equation (5.1) starting at x at time $t = s$. The mappings $\varphi_{s,t}(x)$ are called the stochastic flow of the diffusion process (5.1).

5.2.1 The Ito formula

In this notation, if T_s stands for the first jump time after time s, we have

$$\mathbb{P}\left(T^{(s)} \in dt \ , X_{T^{(s)}} \in dy \mid X_s, \ \varphi_{s,r}(X_s), r \geq s\right)$$

$$= \underbrace{\lambda_t(\varphi_{s,t}(X_s)) \ \exp\left(-\int_s^t \lambda_r(\varphi_{s,r}(X_s)) \ dr\right) 1_{s \leq t} \ dt}_{\mathbb{P}\left(T^{(s)} \in dt \mid X_s, \ \varphi_{s,r}(X_s), \ r \geq s\right)} \times \underbrace{S_t(\varphi_{s,t}(X_s), dy)}_{\mathbb{P}\left(X_{T^{(s)}} \in dy \mid T^{(s)} = t, \ \varphi_{s,r}(X_s), \ r \geq s\right)}$$

$$\Longrightarrow \ \mathbb{P}\left(T^{(t)} \in dt \ , X_{t+dt} \in dy \mid X_t\right) = \lambda_t(X_t) \ dt \ S_t(X_t, dy) \tag{5.6}$$

Expanding a formal Taylor series, for smooth functions f on $\mathbb{R}_+ \times \mathbb{R}^d$ we have

$$\begin{aligned} df(t, X_t) &= f(t + dt, X_t + dX_t) - f(t, X_t) \\ &= \frac{\partial f}{\partial t}(t, X_t) dt \ + \ \sum_{i=1}^d \partial_i f(X_t) \ dX_t^i + \frac{1}{2} \ \sum_{i,j=1}^d \ \partial_{i,j} f(X_t) \ dX_t^i dX_t^j \\ &\quad + \Delta f(t, X_t) - \left[\sum_{i=1}^d \partial_i f(X_t) \ \Delta X_t^i + \frac{1}{2} \ \sum_{i,j=1}^d \ \partial_{i,j} f(X_t) \ \Delta X_t^i \Delta X_t^j\right] \end{aligned} \tag{5.7}$$

The terminology smooth is used to described functions which can be derived at any order and have bounded derivatives with compact support. These regularity conditions can be relaxed using appropriate domains of definitions; we have chosen here these strong conditions to clarify the presentation and concentrate on the stochastic modeling rather than on sophisticated analytical considerations.

The r.h.s. term comes from the fact that

$$dX_t^i = dX_t^{c,i} + \Delta X_t^i$$

with the jump term ΔX_t^i, and the increment

$$dX_t^{c,i} := b_t^i(X_t) \ dt + \sum_{1 \leq j \leq d} \sigma_{j,t}^i(X_t) \ dW_t^j$$

of the continuous process

$$X_t^{c,i} := X_0^i + \int_0^t b_t^i(X_s) \ ds + \sum_{1 \leq j \leq d} \int_0^t \ \sigma_{j,s}^i(X_t) \ dW_s^j$$

In this situation, we have

$$dX_t^{c,i} dX_t^{c,j} = \sum_{1 \leq k,l \leq d} \sigma_{k,t}^i(X_t)\ \sigma_{l,t}^j(X_t)\ dW_t^k dW_t^l$$

and applying the rules $dW_t^k \times dW_t^l = 1_{k=l}\ dt$, and $dt \times dW_t^i = 0$, we conclude that

$$dX_t^{ci} dX_t^{cj} = \sum_{1 \leq k \leq d} \sigma_{k,t}^i(X_t)\ \sigma_{k,t}^j(X_t)\ dt = \left(\sigma_t(\sigma_t)^T\right)_j^i (X_t)\ dt$$

In this context, the quadratic term $dX_t^i dX_t^j$ is interpreted as the increment of the covariation process

$$\begin{aligned} dX_t^i dX_t^j &= \left(\sigma_t(\sigma_t)^T\right)_{i,j} (X_t)\ dt + \Delta X_t^i \Delta X_t^j \\ &:= d\left[X^i, X^j\right]_t = d\langle X^{c,i}, X^{c,j}\rangle_t + \Delta X_t^i \Delta X_t^j \end{aligned}$$

with

$$d\langle X^{c,i}, X^{c,j}\rangle_t = dX_t^{c,i} dX_t^{c,j} = \left(\sigma_t(\sigma_t)^T\right)_j^i (X_t)\ dt$$

Rewritten in terms of the continuous and jump parts of the process, the Doeblin-Itō formula (5.7) takes the form

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t}(t, X_t) dt + \sum_{i=1}^d \partial_i f(X_t)\ dX_t^{c,i} \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \partial_{i,j} f(X_t)\ dX_t^{c,i} dX_t^{c,j} + \Delta f(t, X_t) \end{aligned}$$

To take the final step, we use (5.6) to check that

$$\Delta f(t, X_t) = \mathbb{E}\left(\Delta f(t, X_t) \mid \mathcal{F}_t\right) + \Delta f(t, X_t) - \mathbb{E}\left(\Delta f(t, X_t) \mid \mathcal{F}_t\right)$$

with the predictable part of the increment

$$\mathbb{E}\left(\Delta f(t, X_t) \mid \mathcal{F}_t\right) = \lambda_t(X_t)\ dt \int \left(f(t, y) - f(t, x)\right)\ S_t(x, dy)$$

and the martingale part of the increment

$$dM_t^d(f) = \Delta f(t, X_t) - \mathbb{E}\left(\Delta f(t, X_t) \mid \mathcal{F}_t\right)$$

 This shows that

$$df(t,X_t) = \left[\frac{\partial}{\partial t} + L_t\right] f(t,X_t)\ dt + dM_t(f) \tag{5.8}$$

with the sum $L_t = L_t^c + L_t^d$ of the generators

$$\begin{aligned} L_t^c(f) &= \sum_{i=1}^d a_t^i\ \partial_i f + \frac{1}{2}\ \sum_{i,j=1}^d \left(\sigma_t(\sigma_t)^T\right)_j^i\ \partial_{i,j} f \\ &:= a_t\ \nabla f + \frac{1}{2}\ \mathrm{Tr}\left(\sigma_t\sigma_t^T\nabla^2 f\right) \qquad \text{and} \qquad L_t^d = \lambda_t\ [S_t - Id] \end{aligned} \tag{5.9}$$

and the martingale increments

$$dM_t(f) = dM_t^c(f) + dM_t^d(f) \quad \text{with} \quad dM_t^c(f) := \sum_{i,j=1}^d \partial_i f(X_t)\ \sigma_{j,t}^i(X_t)\ dW_t^j$$

We recall that the angle bracket $\langle \mathcal{M}\rangle_t$ (a.k.a. the predictable quadratic variation) of a given martingale $\mathcal{M}_t$ w.r.t. some filtration $\mathcal{F}_t$ is the predictable stochastic process $\langle \mathcal{M}\rangle_t$ s.t. $\mathcal{M}_t^2 - \langle \mathcal{M}\rangle_t$ is a martingale. By construction, arguing as in the discrete time case, the angle bracket of $M_t(f)$ w.r.t. the filtration $\mathcal{F}_t = \sigma(X_s,\ s \le t)$ is the sum

$$\langle M(f)\rangle_t = \langle M^c(f)\rangle_t + \langle M^d(f)\rangle_t$$

of the angle brackets of the martingales $M_t^c(f)$ and $M_t^d(f)$ given by

$$\langle M^c(f)\rangle_t = \sum_{i,j=1}^d \int_0^t \mathbb{E}\left((dM_s^c(f))^2 \mid \mathcal{F}_s\right) = \sum_{j=1}^d \int_0^t \left(\sum_{i=1}^d \partial_i f(X_s)\ \sigma_{j,s}^i(X_s)\right)^2 ds \tag{5.10}$$

and

$$\begin{aligned} \langle M^d(f)\rangle_t &= \int_0^t \mathbb{E}\left((dM_s^d(f))^2 \mid \mathcal{F}_s\right) = \int_0^t \mathbb{E}\left((\Delta M_s^d(f))^2 \mid \mathcal{F}_s\right) \\ &= \int_0^t \mathbb{E}\left((\Delta f(s,X_s))^2 \mid \mathcal{F}_s\right) = \int_0^t \lambda_s(X_s)\ \left[\int (f(s,y) - f(s,X_s))^2\ S_s(X_s,dy)\right] ds \end{aligned} \tag{5.11}$$

 These angle brackets can be rewritten in a more synthetic form

$$\langle M(f)\rangle_t = \int_0^t \Gamma_{L_s}(f(s,.),f(s,.))(X_s)\ ds$$

in terms of the "carré du champ" operators

$$\begin{aligned} \Gamma_{L_t}(f,f)(x) &:= L_t((f - f(x))^2)(x) \\ &= L_t(f^2)(x) - 2f(x)L_t(f)(x) = \Gamma_{L_t^c}(f,f)(x) + \Gamma_{L_t^d}(f,f)(x) \end{aligned}$$

The "carré du champ" is non negative since we have

$$\Gamma_{L_t}(f,f)(x) \simeq_{h\downarrow 0} \frac{[P_{t,t+h} - Id][(f - f(x))^2]}{h}(x) = \frac{P_{t,t+h}[(f - f(x))^2]}{h}(x) \ge 0$$

We can also show that

$$M_t(f)M_t(g) - \langle M(f), M(g)\rangle_t \quad \text{is a martingale}$$

with the angle bracket

$$d\langle M(f), M(g)\rangle_t = \mathbb{E}\left(dM_t(f)dM_t(g) \mid X_t\right) = \Gamma_L(f,g)(X_t)\ dt$$

defined in terms of the "carré du champ" operators

$$\Gamma_{L_t}(f,g)(x) = L_t((f-f(x))(g-g(x)))(x) = \simeq_{h\downarrow 0} \frac{[P_{t,t+h} - Id][(f-f(x))(g-g(x))]}{h}(x)$$

5.2.2 A local characterization

In this section, we present an alternative, and more intuitive, local characterization of the infinitesimal generator, and its "carré du champ" operator in terms of limiting predictable averages of the increments of $f_t(X_t)$ and the ones of the martingale $M_t(f)$.

Firstly, using (5.8), we readily show that

$$\frac{1}{t-s}\ \mathbb{E}\left(f_t(X_t) - f_s(X_s)\ \mid \mathcal{F}_s\right)$$

$$= \frac{1}{t-s}\ \mathbb{E}\left(\int_s^t \left(\frac{\partial}{\partial \tau} + L_\tau\right)(f_\tau)(X_\tau)\ d\tau \mid \mathcal{F}_s\right)$$

This implies that

$$\frac{1}{t-s}\ \mathbb{E}\left(\ [f_t(X_t) - f_s(X_s)]\ \mid X_s = x\right) \longrightarrow_{t\to s} \left(\frac{\partial}{\partial s} + L_s\right)(f_s)(x)$$

On the other hand, using the fact that $M_t(f)$ and $M_t(f)^2 - \langle M(f)\rangle_t$ are $\mathcal{F}$-martingales, we prove that

$$\mathbb{E}\left([M_t(f) - M_s(f)]^2 \mid \mathcal{F}_s\right)$$

$$= \mathbb{E}\left(M_t(f)^2 \mid \mathcal{F}_s\right) - M_s(f)^2$$

$$= \underbrace{\mathbb{E}\left(M_t(f)^2 - \langle M(f)\rangle_t \mid \mathcal{F}_s\right) - \left(M_s(f)^2 - \langle M(f)\rangle_s\right)}_{=0} + \mathbb{E}\left(\langle M(f)\rangle_t - \langle M(f)\rangle_s \mid \mathcal{F}_s\right)$$

$$= \mathbb{E}\left(\langle M(f)\rangle_t - \langle M(f)\rangle_s \mid \mathcal{F}_s\right)$$

from which we conclude that

$$\mathbb{E}\left([M_t(f) - M_s(f)]^2 \mid \mathcal{F}_s\right) = \mathbb{E}\left(\int_s^t \Gamma_{L_\tau}\left(f_\tau, f_\tau\right)(X_\tau)\ d\tau \mid X_s\right)$$

Arguing as above, we prove that

$$\frac{1}{t-s}\ \mathbb{E}\left([M_t(f) - M_s(f)]^2\ \mid \mathcal{F}_s\right) \longrightarrow_{t\to s} \Gamma_{L_s}\left(f_s, f_s\right)(X_s)$$

or alternatively

$$\frac{1}{t-s}\,\mathbb{E}\left(\left[f_t(X_t)-f_s(X_s)-\int_s^t\left(\frac{\partial}{\partial\tau}+L_\tau\right)(f_\tau)(X_\tau)\,d\tau\right]^2\;\middle|\;X_s=x\right)$$

$$\longrightarrow_{t\to s}\Gamma_{L_s}(f_s,f_s)(x)$$

We end this section, we an informal derivation of the infinitesimal generator of the jump-diffusion process (5.3) in terms of the Markov chain with elementary transitions K_n.

Infinitesimal generators play a central role in stochastic analysis. They provide a complete analytic description of continuous time Markov process in terms of integro-differential operators L'_t acting on some domain $\mathcal{D}$ of sufficiently regular functions f with the following formulae

$$\lim_{s\to t}\left[\frac{P'_{s,t}-Id}{\epsilon}\right](f)(x)=\lim_{s\to t}\epsilon^{-1}\mathbb{E}\left((f(X_t)-f(x))\;\middle|\;X_s=x\right)=L'_t(f)(x)$$

This yields the first order expansion of the semigroup

$$P'_{s,t}=Id+(t-s)\;L'_t\;+\mathrm{O}((t-s))$$

The regularity property of the functions depends on the dynamics of the Markov process under study. The generator of pure jump models are defined on every bounded measurable functions. The one of $\mathbb{R}^d$-valued diffusion processes is a second oder differential operator, only defined on twice differentiable and bounded functions.

Taking $\Delta\simeq 0$ in (5.5), we find the approximations

$$J_n-Id=\left(1-\exp\left(-\Delta\lambda'_{t_n}\right)\right)\;(P'_{t_n}-Id)\simeq\lambda'_{t_n}\;(P'_{t_n}-Id)\;\Delta$$

In much the same way, for any smooth functions f we have

$$\begin{aligned}(M_n-Id)(f)(x)&\simeq\mathbb{E}\left(f\left(a'_{t_{n-1}}(x)\;\Delta+\sigma'_{t_{n-1}}(x)\;\left(W_{t_n}-W_{t_{n-1}}\right)\right)-f(x)\right)\\&\simeq L'_{t_{n-1}}(f)(x)\;\Delta\end{aligned}$$

with the operator L'_t defined by the following formula

$$L'_t\;:=\;\sum_{i=1}^d a'_{t,i}\;\partial_i+\frac{1}{2}\sum_{i,j=1}^d\left(\sigma'_t(\sigma'_t)^T\right)_{i,j}\;\partial_{i,j}\tag{5.12}$$

Combining these couple of approximations, we find that

$$\begin{aligned}(K_n-Id)(f)(x)&=[(Id+(M_n-Id))(Id+(J_n-Id))-Id](f)(x)&(5.13)\\&\simeq(M_n-Id)(f)(x)+(J_n-Id)(f)(x)=L_{t_{n-1}}(f)(x)\;\Delta&(5.14)\end{aligned}$$

with the infinitesimal generator of the jump-diffusion model

$$L_t(f)(x):=L'_t(f)(x)\;+\lambda'_t(x)\int[f(y)-f(x)]\;P'_t(x,dy)\tag{5.15}$$

For a more thorough and rigorous discussion on these probabilistic models, we refer the reader to the seminal books by Daniel Revuz and Marc Yor [506], Ioannis Karatzas and Steven Shreve [372], as well as the book by Stewart Ethier and Thomas Kurtz [392], the one by Bernt Øksendal [471], and my review article with N. Hadjiconstantinou [182].

5.3 Path space and twisted processes

Continuous time stochastic processes are defined in terms of a sequence of random variables X_t with a time index t taking values in the uncountable continuous time axis $\mathbb{R}_+$. As a result, proving the existence and the uniqueness of their distribution on the set of trajectories require some sophisticated probabilistic and analytic tools. In contrast with discrete generation stochastic processes, we don't have any explicit description of these probability measures.

Nevertheless, apart from some mathematical technicalities, the analysis of these path space measures often follows the same construction as in the discrete time case. Furthermore, despite the formal derivation of our constructions, all the formulae presented in this section are mathematically correct.

Path space measures are not only of pure mathematical interest. They are commonly used in engineering sciences, statistical machine learning, reliability analysis, as well as in mathematical finance.s Most of these applications rely on Bayesian inference, maximum likelihood estimation, importance sampling techniques and Girsanov type change of probability measures.

To guide the reader's intuition we present a rather informal discussion on these path space models and their application models.

5.3.1 Pure jump models

Path space measures

We start with a pure jump process X_t with intensity function $\lambda_t(x)$ and jump amplitude transition $M_t(x, dy)$ on some state space S. In other words, X_t is a stochastic process with infinitesimal generator

$$L_t(f)(x) = \lambda_t(x) \int [f(y) - f(x)] \, M_t(x, dy)$$

We let $(T_k)_{k\geq 0}$ be the sequence of jump times of X_t, with the convention $T_0 = 0$, and we assume that $X_0 = \omega_0$ for some $\omega_0 \in S$.

The random trajectories of the process on some time interval $[0, t]$ are càdlag paths

$$\omega \ : \ s \in [0, t] \ \mapsto \ \omega_s \in S$$

with a finite number, say n, of jump epochs $t_k \in [0, t]$, $k \leq n$, defined by the fact that $\Delta\omega_{t_k} \neq 0$, with $k \leq n$. In addition, between the jumps the trajectory remains constant, in the sense that $\omega_s = \omega_{t_k}$, for any $s \in [t_k, t_{k-1}[$. We let $D([0, t], S)$ be the set of all the càdlag paths from $[0, t]$ into S, and $D_0([0, t], S)$ the subset of the càdlag piecewise constant paths, with a finite number of jump times.

By construction, we have

$$\mathbb{P}\left((T_{k+1}, X_{T_{k+1}}) \in d(t_{k+1}, \omega_{t_{k+1}}) \ \mid (T_k, X_{T_k}) \in d(t_k, \omega_{t_k})\right)$$

$$= \lambda_{t_{k+1}}(\omega_{t_k}) \ \exp\left(-\int_{t_k}^{t_{k+1}} \lambda_s(\omega_{t_k}) \, ds\right) \, dt_{k+1} \times M_{t_{k+1}}(\omega_{t_k}, d\omega_{t_{k+1}})$$

and

$$\mathbb{P}\left(T_{n+1} \geq t \ \mid (T_n, X_{T_n}) \in d(t_n, \omega_{t_n})\right) = \exp\left(-\int_{t_n}^{t} \lambda_s(\omega_{t_n}) \, ds\right)$$

In the above displayed formula, dt_k and $d\omega_{t_k}$ stands for an infinitesimal neighborhood of the points $t_k \in [0, t]$ and $\omega_{t_k} \in S$.

In the further development of this section, $(t_k)_{1\leq k\leq n} \in [0, t]^n$ stands for the jump times or a trajectory $\omega = (\omega_s)_{s\leq t} \in D_0([0, t], S)$. Recalling that $\omega_s = \omega_{t_k}$, for any $s \in [t_k, t_{k+1}[$, we have the following construction.

The path space distribution of $X = (X_s)_{s\leq t}$ is defined for any $\omega = (\omega_s)_{s\leq t} \in D_0([0,t], S)$ by

$$\begin{aligned}
&\mathbb{P}\left(X \in d\omega\right) \\
&:= \mathbb{P}\left((T_1, X_{T_1}) \in d(t_1, \omega_{t_1}), \ldots, (T_n, X_{T_n}) \in d(t_n, \omega_{t_n}) \ , T_{n+1} \geq t\right) \\
&= \exp\left(-\int_0^t \lambda_s(\omega_{s-})\ ds\right) \times \prod_{s\leq t\ :\ \Delta\omega_s \neq 0} \left[\lambda_s(\omega_{s-})\ ds\ M_s(\omega_{s-}, d\omega_s)\right]
\end{aligned} \tag{5.16}$$

One direct consequence of this result is the conditional distribution formula

$$\begin{aligned}
&\mathbb{P}\left((X_r)_{s\leq r\leq t} \in d(\omega_s)_{s\leq r\leq t}) \ \mid (X_r)_{0\leq r<s} = (\omega_r)_{0\leq r<s}\right) \\
&= \mathbb{P}\left((X_r)_{s\leq r\leq t} \in d(\omega_s)_{s\leq r\leq t}) \ \mid X_{s-} = \omega_{s-}\right) \\
&= \exp\left(-\int_s^t \lambda_r(\omega_{r-})\ dr\right) \times \prod_{s\leq r\leq t\ :\ \Delta\omega_r \neq 0} \left[\lambda_r(\omega_{r-})\ dr\ M_r(\omega_{r-}, d\omega_r)\right]
\end{aligned}$$

There exists a unique path space measure $\mathbb{P}$ on $D([0,\infty[, \mathbb{R})$ s.t. its restrictions to every time mesh sequence are given by (5.20). This path space measure is the distribution of the pure jump process $X = (X_s)_{s\in[0,t]}$.

Proof :
One strategy is to consider the set of paths sequence indexed by rationals and taking values in the compact space $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$:

$$\Omega := D([0,\infty[\cap\ \mathbb{Q}, \mathbb{R}) := \prod_{t\in\mathbb{Q}} \overline{\mathbb{R}} = \left\{(\omega_s)_{s\in\mathbb{Q}} \ :\ \omega_s \in \overline{\mathbb{R}} \ \ \forall s \in \mathbb{Q}\right\}$$

By construction the product space Ω is a compact metrizable state space. Then, we interpret the measures $\mathbb{P}$ defined in (5.16) as positive linear functionals on the space $C(\Omega)$ of real valued continuous functions over $\Omega_{\mathbb{Q}}$ s.t. $\mathbb{P}(1) = 1$. For any $F \in C(\Omega)$ that only depends on the values of ω on some mesh sequence $t_n \in \mathbb{Q}$ we have

$$\mathbb{P}(F) = \int\ F(\omega)\ \mathbb{P}\left(X \in d\omega\right)$$

with the measure $\mathbb{P}$ defined in (5.20). This proves the existence and the consistency of these measures on the subspace $C_{finite}(\Omega)$ of functions that only depends on the values of ω on some finite mesh sequence $t_n \in \mathbb{Q}$. Since $C_{finite}(\Omega)$ is dense in $C(\Omega)$, invoking the Stone-Weierstrass theorem, we conclude that there exists a unique extension to the set $C(\Omega)$. This ends the proof of the theorem. ■

The Poisson process with time non homogenous intensity λ_t corresponds to the situation where $M_t(x, dy) = \delta_{x+1}(dy)$ and $\lambda_t(x) = \lambda_t$.

In this situation, for any $\omega \in D_0([0,t], S)$, such that $\Delta\omega_{t_k} = 1$, with $k \leq n$, we have

$$\begin{aligned}
\mathbb{P}\left(X \in d\omega\right) &= \exp\left(-\int_0^t \lambda_s\ ds\right) \times \prod_{s\leq t\ :\ \Delta\omega_s\neq 0} \left[\lambda_s\ ds\right] \\
&= \exp\left(-\int_0^t \lambda_s\ ds + \int_0^t \log\left(\lambda_s\right)\ d\omega_s\right)\ dt_1 \ldots dt_n
\end{aligned} \tag{5.17}$$

The corresponding path space measure on $\omega \in D_0([0,t],S)$ is called the Poisson measure.
For time homogenous models, this formula also reduces to

$$\begin{aligned}\mathbb{P}(X \in d\omega) &= e^{-\lambda t}\ \lambda^n\ dt_1 \dots dt_n \\ &= e^{-\lambda t}\ \lambda^{\omega_t}\ dt_1 \dots dt_n\end{aligned}$$

Likelihood functionals

We let Λ be some positive valued random variable. Given $\Lambda = \lambda$, we let $X = (X_s)_{s \le t}$ be a Poisson process with intensity λ. Combining

$$\mathbb{P}(X \in d\omega \mid \Lambda = \lambda) = e^{-\lambda t}\ \lambda^{\omega_t}\ dt_1 \dots dt_n$$

with Bayes' rule, we find that the conditional distribution of Λ given a realization

$$X = (X_s)_{s \le t} = (\omega_s)_{s \le t} = \omega$$

of the Poisson process is given by the formula

$$\mathbb{P}[\Lambda \in d\lambda \mid X = \omega] \propto \exp(L(\lambda \mid \omega_t)) \times \mathbb{P}(\Lambda \in d\lambda)$$

with the log likelihood function

$$L(\lambda \mid \omega_t) := -\lambda t\ +\ \omega_t\ \log \lambda$$

We observe that the maximum value of the log-likelihood function is given by

$$\frac{\partial}{\partial \lambda} L(\lambda \mid \omega_t) = -t\ +\ \omega_t\ \frac{1}{\lambda} = 0 \Leftrightarrow \lambda = \omega_t / t$$

We let $\Theta = \left(\Theta^1, \Theta^2\right)$ be some positive valued random variables. Given $\Theta = \theta = (\theta^1, \theta^2)$, we let $X = (X_s)_{s \le t}$ be a time non homogeneous Poisson process with power law intensity function $\lambda_t = \theta^1 \theta^2 t^{\theta_2 - 1}$. In this situation, we have

$$\begin{aligned}&\mathbb{P}(X \in d\omega \mid \Lambda = \lambda) \\ &= \exp\left(-\int_0^t \lambda_s\ ds\right) \times \left[\prod_{1 \le k \le n} \lambda_{t_k}\right]\ dt_1 \dots dt_n \\ &= \exp\left(-\theta^1 t^{\theta_2}\right) \times (\theta^1 \theta^2)^n\ \left[\prod_{1 \le k \le n} t_k^{\theta_2 - 1}\right]\ dt_1 \dots dt_n\end{aligned}$$

Using Bayes' rule, we find that the conditional distribution of Θ given a realization $(X_s)_{s \le t} = (\omega_s)_{s \le t}$ of the Poisson process is given by

$$\begin{aligned}&\mathbb{P}[\Theta \in d\theta \mid X = \omega] \\ &\propto \exp[L(\theta \mid \omega)] \times \mathbb{P}(\Theta \in d\theta)\end{aligned}$$

with the log likelihood function

$$L(\theta \mid \omega) := -\theta^1 t^{\theta_2} + \omega_t \ \log\left(\theta^1\theta^2\right) + (\theta_2 - 1) \sum_{1\leq k\leq \omega_t} \log t_k$$

We observe that the maximum value of the log-likelihood function is given by the point θ s.t.

$$\left\{\begin{array}{rcl} \frac{\partial}{\partial\theta^1} L(\theta \mid (\omega_s)_{s\leq t}) & = & -t^{\theta^2} + \frac{\omega_t}{\theta^1} = 0 \iff \omega_t = \theta^1 t^{\theta^2} \\ & = & \\ \frac{\partial}{\partial\theta^2} L(\theta \mid (\omega_s)_{s\leq t}) & = & -\underbrace{\theta^1 \ t^{\theta^2}}_{=\omega_t} \ \log t + \frac{\omega_t}{\theta^2} + \log\left(\prod_{1\leq k\leq n} t_k\right) = 0 \end{array}\right.$$

This implies that

$$\theta^2 = \omega_t \ \left(\log \frac{t^{\omega_t}}{\prod_{1\leq k\leq n} t_k}\right)^{-1} = \omega_t \ \left(\sum_{1\leq k\leq n} \log \frac{t}{t_k}\right)^{-1}$$

and

$$\theta^1 = \omega_t \ \exp\left(-\frac{1}{1 - \frac{1}{\omega_t}\sum_{1\leq k\leq \omega_t} \frac{\log t_k}{\log t}}\right)$$

Girsanov's transformations

We let (X_t, X'_t) be a pair of pure jump process with *positive* intensity functions $(\lambda_t(x), \lambda'_t(x))$, and jump amplitude transitions $(M_t(x, dy), M'_t(x, dy))$ on some state space S.

We further assume that M_t and M'_t have a density m_t and m'_t with respect to some measure ν on S; that is, we have that

$$M_t(x, dy) = m_t(x, y)\ \nu(dy) \quad \text{and} \quad M'_t(x, dy) = m'_t(x, y)\ \nu(dy)$$

For instance, all the Gaussian transitions on $S = \mathbb{R}$ are absolutely continuous w.r.t. the Lebesgue measure $\nu(dx) = dx$.

In this situation, the path space measures of

$$X = (X_s)_{s\leq t} \quad \text{and} \quad X' = (X'_s)_{s\leq t}$$

are defined for any $\omega = (\omega_s)_{s\leq t} \in D_0([0, t], S)$ (with jump times $(t_k)_{1\leq k\leq n} \in [0, t]^n$) by the formulae

$$\begin{aligned} &\mathbb{P}(X \in d\omega) \\ &= \exp\left(-\int_0^t \lambda_s(\omega_{s-})\ ds\right) \times \prod_{s\leq t\ :\ \Delta\omega_s\neq 0} \left[\lambda_s(\omega_{s-})\ ds\ m_s(\omega_{s-}, \omega_s)\ \nu(d\omega_s)\right] \end{aligned}$$

and

$$\begin{aligned} &\mathbb{P}\left(X' \in d\omega\right) \\ &= \exp\left(-\int_0^t \lambda'_s(\omega_{s-})\ ds\right) \times \prod_{s\leq t\ :\ \Delta\omega_s\neq 0} \left[\lambda'_s(\omega_{s-})\ ds\ m'_s(\omega_{s-}, \omega_s)\ \nu(d\omega_s)\right] \end{aligned}$$

It is readily check that

$$\mathbb{P}\left(X' \in d\omega\right) = \mathcal{Z}(\omega) \times \mathbb{P}\left(X \in d\omega\right)$$

with the function

$$\mathcal{Z}(\omega) = \exp\left(-\int_0^t \left[\lambda'_s - \lambda_s\right](\omega_{s-})\ ds\right) \times \prod_{s\leq t\ :\ \Delta\omega_s\neq 0} \left[\frac{\lambda'_s(\omega_{s-})}{\lambda_s(\omega_{s-})}\ \frac{m'_s(\omega_{s-},\omega_s)}{m_s(\omega_{s-},\omega_s)}\right]$$

The above formula is called the Girsanov change of measure, or a Girsanov transformation. It is often written for any function F on the path space $D_0([0,t],S)$ as follows

$$\mathbb{E}\left(F\left((X'_s)_{s\leq t}\right)\right) = \mathbb{E}\left(F\left((X_s)_{s\leq t}\right)\ Z_t\right) \tag{5.18}$$

with the exponential stochastic processes Z_t defined by

$$Z_t = \exp\left(-\int_0^t \left[\lambda'_s - \lambda_s\right](X_{s-})\ ds\right) \times \prod_{0\leq s\leq t\ :\ \Delta\omega_s\neq 0} \left[\frac{\lambda'_s(X_{s-})}{\lambda_s(X_{s-})}\ \frac{m'_s(X_{s-},X_s)}{m_s(X_{s-},X_s)}\right]$$

Exponential martingales

The Poisson processes $(X_t, X'_t) = (N_t, N'_t)$ with time non homogenous intensity (λ_t, λ'_t) corresponds to the situation where

$$M_t(x,dy) = M'_t(x,dy) = \delta_{x+1}(dy) \quad \text{and} \quad (\lambda_t(x), \lambda'_t(x)) = (\lambda_t, \lambda'_t)$$

In this situation, for any $\omega \in D_0([0,t],S)$, such that $\Delta\omega_{t_k} = 1$, with $k \leq n$, we have

$$\mathcal{Z}(\omega) \ = \ \exp\left[-\int_0^t \left[\lambda'_s - \lambda_s\right]\ ds + \int_0^t \log\left(\lambda'_s/\lambda_s\right)\ d\omega_s\right]$$

Sometimes, we also rewrite this function in the following form

$$\mathcal{Z}(\omega) = \exp\left(-\int_0^t \left[\lambda'_s - \lambda_s\right]\ ds\right) \prod_{0\leq s\leq t} \left(1 + \left(\frac{\lambda'_s}{\lambda_s} - 1\right)\ \Delta\omega_s\right)$$

We also have that

$$\begin{aligned} Z_t &= \exp\left[-\int_0^t \left[\lambda'_s - \lambda_s\right]\ ds + \int_0^t \log\left(\lambda'_s/\lambda_s\right)\ dN_s\right] \\ &= \exp\left(-\int_0^t \left[\lambda'_s - \lambda_s\right]\ ds\right) \prod_{0\leq s\leq t} \left(1 + \left(\frac{\lambda'_s}{\lambda_s} - 1\right)\ \Delta N_s\right) \end{aligned}$$

In terms of the martingale

$$dM^\lambda_t = \left(\frac{\lambda'_t}{\lambda_t} - 1\right)(dN_t - \lambda_t dt) = \underbrace{\left(\frac{\lambda'_t}{\lambda_t} - 1\right)\ dN_t}_{\Delta M^\lambda_t} - \left(\lambda'_t - \lambda_t\right)\ dt$$

we also have the exponential formula

$$Z_t = e^{M_t^\lambda} \prod_{0\leq s\leq t} \left(\left(1+\Delta M_s^\lambda\right)\ e^{-\Delta M_s^\lambda} \right)$$

We also notice that

$$\begin{aligned} Z_{t+dt} - Z_t &= Z_t \left(e^{\left(M_{t+dt}^\lambda - M_t^\lambda\right) - \Delta M_t^\lambda} \left(1+\Delta M_t^\lambda\right) - 1 \right) \\ &= Z_t \left(e^{-(\lambda_t' - \lambda_t)dt} \left[1+\Delta M_t^\lambda\right] \ - 1 \right) \\ &= Z_t \left(\left[1-\left(\lambda_t' - \lambda_t\right) dt\right] \ \left[1+\Delta M_t^\lambda\right] - 1 \right) \\ &= Z_t \left(\Delta M_t^\lambda - \left(\lambda_t' - \lambda_t\right) dt \right) = Z_t\ dM_t^\lambda \end{aligned}$$

This implies that

$$dZ_t = Z_t\ dM_t^\lambda$$

In this notation, the change of measure formula (5.18) takes the following form.

For any $t \geq 0$, and any functional F on $D_0([0,t],\mathbb{R})$ we have

$$\mathbb{E}\left(F\left((N_s)_{s\leq t}\right)\ Z_t\right) = \mathbb{E}\left(F\left((N_s')_{s\leq t}\right)\right) \tag{5.19}$$

with the martingale $dZ_t = Z_t\ dM_t^\lambda$ defined by the exponential formula

$$Z_t = \exp\left(M_t^\lambda\right) \prod_{0\leq s\leq t} \left\{ \left(1+\Delta M_s^\lambda\right)\ \exp\left(-\Delta M_s^\lambda\right) \right\}$$

and the martingale increments

$$dM_t^\lambda = \left(\lambda_t' \lambda_t^{-1} - 1\right)\left(dN_t - \lambda_t\ dt\right)$$

If we interpret $dZ_t = Z_t - Z_{t-dt}$, then it is preferable to write $dZ_t = Z_{t-}\ dM_t^\lambda$. In our notational system, we interpret $dZ_t = Z_{t+dt} - Z_t$ so that $dZ_t = Z_t\ dM_t^\lambda$ is well defined infinitesimal increment.

For time homogenous models $(\lambda_t, \lambda_t') = (\lambda, \lambda')$, this formula also reduces to

$$\mathcal{Z}(\omega) \ = \ e^{(\lambda-\lambda')t} \times \left(\lambda'/\lambda\right)^{\omega_t} \iff Z_t = e^{(\lambda-\lambda')t} \times \left(\lambda'/\lambda\right)^{N_t}$$

5.3.2 Diffusion models

The Wiener measure

The random trajectories of a Brownian process on some time interval $[0,t]$ are continuous paths

$$\omega \ : \ s\in[0,t] \ \mapsto \ \omega_s \in \mathbb{R}$$

We let $C([0,t],S)$ be the set the continuous trajectories from $[0,t]$ into $\mathbb{R}$.

We consider a time mesh sequence $t_n = nh$, with $n \in \mathbb{N}$, and a given time step $\Delta t := h \in]0,1[$. By construction, the path space distribution of $W = (W_{t_k})_{0\leq k\leq n}$ is defined for any $\omega = (\omega_s)_{s\leq t_n} \in C([0,t_n],\mathbb{R})$ by

$$\begin{aligned} \mathbb{P}\left(W \in d\omega\right) &:= \mathbb{P}\left(W_{t_1} \in d\omega_{t_1}, \ldots,\ W_{t_n} \in d\omega_{t_n}\right) \\ &= \frac{1}{(2\pi\Delta t)^{n/2}}\ \exp\left[-\frac{1}{2} \sum_{1\leq k\leq n} \left(\frac{\Delta\omega_{t_k}}{\Delta t}\right)^2\ \Delta t \right] d\omega_{t_1} \ldots d\omega_{t_n} \end{aligned} \tag{5.20}$$

Since the path space measure of $W := (W_s)_{s\leq t}$ is supported by the set of functions $\omega = (\omega_s)_{s\leq t}$ that are nowhere differentiable we cannot pass in the limit $\Delta t \to 0$ in the above formula. Nevertheless, using the same lines of arguments as in the ones we used on page 95, we have the following result.

There exists a unique path space measure $\mathbb{P}$ on $C([0,\infty[,\mathbb{R})$ s.t. its restrictions to every time mesh sequence are given by (5.20) This path space measure is the distribution of the Brownian motion $W = (W_s)_{s\geq 0}$, and it is called the Wiener measure on the set of continuous trajectories $\Omega := C([0,\infty[,\mathbb{R})$. The set Ω equipped with the Wiener measure $\mathbb{P}$ is called the Wiener space.

Path space diffusions

To simplify the presentation, we restrict ourselves with a 1-dimensional diffusion X_t defined by the SDE

$$dX_t = b(X_t)\ dt + dW_t \tag{5.21}$$

with some regular homogenous function b.

We denote by $X^h_{t_n}$ the discrete approximation model on some time mesh defined by

$$\Delta X^h_{t_{n+1}} \ := \ X^h_{t_{n+1}} - X^h_{t_n} = b_{t_n}(X^h_{t_n})\ h + \ \sigma_{t_n}(X^h_{t_n})\ \Delta W_{t_{n+1}} \tag{5.22}$$

We assume that $X^h_0 = \omega_0 = X_0$, for some $\omega_0 \in \mathbb{R}$.

By construction, the the path space distribution of $X^h = (X^h_{t_k})_{0\leq k\leq n}$ is defined for any $\omega = (\omega_s)_{s\leq t_n} \in C([0,t_n],\mathbb{R})$ by

$$\begin{aligned}\mathbb{P}\left(X^h \in d\omega\right) \ &:= \ \mathbb{P}\left(X^h_{t_1} \in d\omega_{t_1}, \ldots,\ X^h_{t_n} \in d\omega_{t_n}\right)\\ &= \ \frac{1}{(2\pi\Delta t)^{n/2}}\ \exp\left(-\frac{1}{2}\sum_{1\leq k\leq n}\ \left[\frac{\Delta\omega_{t_k}}{\Delta t} - b(\omega_{t_{k-1}})\ \right]^2 \Delta t\right)\ d\omega_{t_1}\ldots d\omega_{t_n}\end{aligned}$$

It is now readily checked that

$$\mathbb{P}\left(X^h \in d\omega\right) := \mathcal{Z}^h(\omega) \times \mathbb{P}\left(W \in d\omega\right)$$

with the density function

$$\mathcal{D}^h(\omega) = \exp\left(\sum_{1\leq k\leq n}\ b(\omega_{t_{k-1}})\ \Delta\omega_{t_k} - \frac{1}{2}\sum_{1\leq k\leq n} b^2(\omega_{t_{k-1}})\ \Delta t\right)$$

Choosing $t_n = h\lfloor t/h\rfloor$, and taking the limit $h = \Delta t \to 0$, we obtain the path space measure of the diffusion process $X = (X_s)_{s\leq t}$ on $C([0,t],\mathbb{R})$

$$\mathbb{P}\left(X \in d\omega\right) := D(\omega) \times \mathbb{P}\left(W \in d\omega\right) \tag{5.23}$$

with the density function

$$D(\omega) = \exp\left(\ \int_0^t\ b(\omega_s)\ d\omega_s - \frac{1}{2}\ \int_0^t\ b^2(\omega_s)\ ds\right)$$

Girsanov transformations

The formula (5.23) is called the Cameron Martin formula or the Girsanov's theorem for diffusion processes.

As in the pure jump case, this formula is often rewritten as follows.

For any function F on the path space $C([0,t],\mathbb{R})$ we have

$$\mathbb{E}\left(F\left((X_s)_{s\leq t}\right)\right) = \mathbb{E}\left(F\left((W_s)_{s\leq t}\right)\ D\left((W_s)_{s\leq t}\right)\right) \tag{5.24}$$

with the density function D on the path space $C([0,t],\mathbb{R})$ defined by

$$Z_t := D\left((W_s)_{s\leq t}\right) = \exp\left(\int_0^t\ b(W_s)\ dW_s - \frac{1}{2}\int_0^t\ b^2(W_s)\ ds\right)$$

Notice that

$$Z_t = e^{U_t} \quad \text{with} \quad U_t := \int_0^t\ b(W_s)\ dW_s - \frac{1}{2}\int_0^t\ b^2(W_s)\ ds$$

Combining Doeblin-Itō formula with the fact that

$$dU_t = b(W_t)\ dW_t - \frac{1}{2}b(W_t)^2\ dt \quad \text{and} \quad dU_t dU_t = b(W_t)^2\ dW_t^2 = b(W_t)^2\ dt$$

we prove that Z_t is a martingale with increments given by

$$dZ_t = e^{U_t}\ dU_t + \frac{1}{2}\ e^{U_t}\ dU_t dU_t = Z_t\ b(W_t)\ dW_t$$

Girsanov Theorem:
We let X_t be the diffusion process defined as in (5.21) by replacing $b(x)$ by some regular function $b_t(x)$ and taking $\sigma_t = 1$. For any $t \geq 0$, and any functional F on $C([0,t],\mathbb{R})$ we have

$$\mathbb{E}\left(F\left((X_s)_{s\leq t}\right)\ L_t\right) = \mathbb{E}\left(F\left((W_s)_{s\leq t}\right)\right) \tag{5.25}$$

with the martingale

$$dL_t = L_t\ b(X_t)\ dW_t$$

given by the exponential formula

$$L_t = \exp\left(-\int_0^t\ b(X_s)\ dW_s - \frac{1}{2}\int_0^t\ b^2(X_s)\ ds\right)$$

Proof :
We first observe that

$$\exp\left(-\int_0^t\ b(X_s)\ dW_s - \tfrac{1}{2}\int_0^t\ b^2(X_s)\ ds\right) = \exp\left(-\int_0^t\ b(X_s)\ dX_s + \tfrac{1}{2}\int_0^t\ b^2(X_s)\ ds\right)$$

We check this claim using the fact that

$$\begin{aligned} dX_t = b(X_t)\ dt + dW_t \quad &\Longrightarrow \quad dW_t = dX_t - b(X_t)\ dt \\ &\Longrightarrow \quad b(X_t)\ dW_t = b(X_t)\ dX_t - b^2(X_t)\ dt \end{aligned}$$

Replacing in (5.24) the function $F\left((X_s)_{s\leq t}\right)$ by the function

$$F\left((X_s)_{s\leq t}\right)\ \exp\left(-\int_0^t\ b(X_s)\ dW_s-\int_0^t\ b^2(X_s)\ ds\right)$$

$$=F\left((X_s)_{s\leq t}\right)\exp\left(-\int_0^t\ b(X_s)\ dX_s+\tfrac{1}{2}\int_0^t\ b^2(X_s)\ ds\right)$$

we find that

$$\mathbb{E}\left(F\left((X_s)_{s\leq t}\right)\exp\left(-\int_0^t\ b(X_s)\ dW_s-\tfrac{1}{2}\int_0^t\ b^2(X_s)\ dr\right)\right)$$

$$=\mathbb{E}\left(F\left((X_s)_{s\leq t}\right)\exp\left(-\int_0^t\ b(X_s)\ dX_s+\tfrac{1}{2}\int_0^t\ b^2(X_s)\ ds\right)\right)$$

The end of the proof is now a direct consequence of (5.24). This ends the proof of the theorem. ∎

The Girsanov theorem stated above and the formula (5.23) can be extended to d-dimensional non homogeneous diffusions

$$dX_t=b_t(X_t)\ dt+dW_t \tag{5.26}$$

where W_t is a d-dimensional Brownian motion (i.e. $W_t=\left(W_t^i\right)_{1\leq i\leq d}$ with d independent Brownian motions W_t^i, $1\leq i\leq d$), and b_t is a function taking values in $\mathbb{R}^d$. In this situation, the formulae (5.23), (5.24), and (5.25) are valid with

$$Z_t=\exp\left(\int_0^t\ b_s(W_s)'\ dW_s-\frac{1}{2}\int_0^t\ \|b_s(W_s)\|^2\ ds\right)$$

and

$$L_t=\exp\left(-\int_0^t\ b_s(X_s)'\ dW_s-\frac{1}{2}\int_0^t\ \|b_s(X_s)\|^2\ ds\right)$$

5.3.3 Exponential twisted measures

We let X_t be some Markov process on some state space E with infinitesimal generator L_t.

We also consider a collection of sufficiently smooth positive functions h_t, and we set

$$M_t^h:=h_0^{-1}(X_0)h_t(X_t)\ \exp\left(-\int_0^t\ \left[h_s^{-1}(\partial_s+L_s)h_s\right](X_s)ds\right)$$

with $h_t^{-1}=1/h_t$. Using Doeblin-Itō formula (5.8), we prove that M_t^h is a positive martingale with unit mean.

More precisely, we have

$$\begin{aligned}dM_t^h&=M_t^h\ \left\{\left(h_t^{-1}\left(\partial_t h_t+L_t(h_t)\right)\right)(X_t)\ dt+h_t^{-1}(X_t)dM_t(h)\right\}-M_t^h\ [h_t^{-1}L_t(h_t)](X_t)\ dt\\&=h_t^{-1}(X_t)M_t^h dM_t(h)\end{aligned} \tag{5.27}$$

We let X^h be the process defined by the change of probability measure

$$\mathbb{E}\left(F((X_s^h)_{s\leq t})\ \right)=\mathbb{E}\left(F((X_s)_{s\leq t})\ M_t^h\right)$$

The conditional expectations w.r.t. $(X_r^h)_{r\leq s}$, with $s\leq t$ are given by the formula

$$\mathbb{E}\left(F_1((X_r^h)_{r\leq s})\ F_2((X_r^h)_{s\leq r\leq t})\ \right)\ =\ \mathbb{E}\left(F_1((X_r^h)_{r\leq s})\ \mathbb{E}\left(F_2((X_r^h)_{s\leq r\leq t})\mid(X_r^h)_{r\leq s}\right)\ \right)$$

On the other hand, using the Markov property we have

$$\begin{aligned}
&\mathbb{E}\left(F_1((X^h_r)_{r\le s})\ F_2((X^h_r)_{s\le r\le t})\ \right)\\
&= \mathbb{E}\left(F_1((X_r)_{r\le s})M^h_s\ \mathbb{E}\left(F_2((X_r)_{s\le r\le t})M^h_t/M^h_s \mid X_s\right)\ \right)\\
&= \mathbb{E}\left(F_1((X_r)_{r\le s})M^h_s\ \mathbb{P}^h_{s,X_s}(F_2)\ \right) = \mathbb{E}\left(F_1((X^h_r)_{r\le s})\ \mathbb{P}^h_{s,X^h_s}(F_2)\ \right)
\end{aligned}$$

with the functional

$$\begin{aligned}
\mathbb{P}^h_{s,x}(F_2) &:= \mathbb{E}\left(F_2((X_r)_{s\le r\le t})M^h_t/M^h_s \mid X_s = x\right)\\
&= h_s^{-1}(x)\ \mathbb{E}\left(F_2((X_r)_{s\le r\le t})\ h_t(X_t)\ \exp\left(-\int_s^t h_s^{-1}(X_s)(\partial_s + L_s)(h_s)(X_s)ds\right) \mid X_s = x\right)
\end{aligned}$$

This implies that

$$\mathbb{E}\left(F_2((X^h_r)_{s\le r\le t}) \mid (X^h_r)_{r\le s}\right) = \mathbb{E}\left(F_2((X^h_r)_{s\le r\le t}) \mid X^h_s\right) = \mathbb{P}^h_{s,X^h_s}(F_2)$$

To get one step further, for any sufficiently regular function f on E, combining (5.27) with Doeblin-Itō formula (5.8) we prove that

$$\begin{aligned}
d(f(X_t)M^h_t) &= M^h_t \overbrace{(L_t(f)(X_t)dt + dM_t(f))}^{df(X_t)} + f(X_t)\overbrace{h_t^{-1}(X_t)M^h_t dM_t(h)}^{=dM^h_t}\\
&\quad + \mathbb{E}\left(df(X_t)dM^h_t \mid (X_s)_{s\le t}\right) + \underbrace{df(X_t)dM^h_t - \mathbb{E}\left(df(X_t)dM^h_t \mid (X_s)_{s\le t}\right)}_{:=\mathcal{M}_t}
\end{aligned}$$

for some martingale $\mathcal{M}_t$. On the other hand, we have

$$\begin{aligned}
\mathbb{E}\left(df(X_t)dM^h_t \mid (X_s)_{s\le t}\right) &= \mathbb{E}\left(dM_t(f)dM^h_t \mid (X_s)_{s\le t}\right)\\
&= h_t^{-1}(X_t)M^h_t\ \mathbb{E}\left(dM_t(f)dM(h)_t \mid (X_s)_{s\le t}\right)\\
&= h_t^{-1}(X_t)M^h_t\ d\langle M(h), M(f)\rangle_t = M^h_t\ \left(h_t^{-1}\Gamma_{L_t}(h_t, f)\right)(X_t)\ dt
\end{aligned}$$

This yields

$$\begin{aligned}
f(X_{t+dt})\ M^h_{t+dt}/M^h_t - f(X_t) &= \left(f(X_{t+dt})M^h_{t+dt} - f(X_t)M^h_t\right)/M^h_t\\
&= \left[L_t(f) + h_t^{-1}\Gamma_{L_t}(h_t, f)\right](X_t)dt + d\widetilde{\mathcal{M}}_t
\end{aligned}$$

for some martingale $\widetilde{\mathcal{M}}_t$. In particular, This implies that

$$\begin{aligned}
\frac{1}{dt}\left[\mathbb{E}\left(f(X^h_{t+dt}) \mid X^h_t = x\right) - f(x)\right] &= \frac{1}{dt}\left[\mathbb{E}\left(f(X^h_{t+dt})\ M^h_{t+dt}/M^h_t \mid X_t = x\right) - f(x)\right]\\
&\simeq_{dt\downarrow 0} L^{[h]}_t(f)(x) := L_t(f)(x) + h_t^{-1}(x)\Gamma_{L_t}(h_t, f)(x)
\end{aligned}$$

This show that X^h_t has infinitesimal generator

$$L^{[h]}_t(f) := L_t(f) + h_t^{-1}\Gamma_{L_t}(h_t, f)$$

For the diffusion generator $L_t = L_t^c$ defined in (5.9), using (5.10) we have

$$L_t^{[h]}(f) = \sum_{i=1}^{d} a_t^{h,i}\ \partial_i f$$

with the drift functions $a_t^{h,i}$ defined for any $1 \leq i \leq d$ by

$$a_t^{h,i} := a_t^i + \sum_{k=1}^{d} \left(\sigma_t \sigma_t^T\right)_i^k\ \partial_k \log h_t \qquad \left(\text{with} \quad \left(\sigma_t \sigma_t^T\right)_i^k = \sum_{j=1}^{d} \sigma_{j,t}^k \sigma_{j,t}^i\right)$$

For the jump generator $L_t = L_t^d$ defined in (5.9), using (5.11) we find that

$$\begin{aligned} L_t^{[h]}(f)(x) &= L_t(f)(x) + \lambda_t(x) \int (f(y) - f(x)) \left(\frac{h_t(y)}{h_t(x)} - 1\right) S_t(x, dy) \\ &= \lambda_t^h(x) \int (f(y) - f(x))\ S_t^h(x, dy) \end{aligned}$$

with the jump intensity and transition kernels

$$\lambda_t^h := \lambda_t\ h_t^{-1}\ S_t(h) \quad \text{and} \quad S_t^h(x, dy) := \frac{S_t(x, dy)\ h_t(y)}{S_t(h)(x)}$$

5.4 Evolution semigroups

We return to the jump-diffusion process X_t discussed in section 5.2. Its conditional transitions are given for any $0 \leq r \leq s \leq t$, and any bounded Borel function f on $\mathbb{R}^d$ by the formulae

$$\begin{aligned} P_{r,t}(f)(x) &:= \mathbb{E}\left(f(X_t) \mid X_r = x\right) = \mathbb{E}\left(\varphi_{0,t}\left(x\right)\right) \\ &= = \mathbb{E}\left(\mathbb{E}\left(f(X_t) \mid X_s\right) \mid X_r = x\right) = P_{r,s}(P_{s,t}(f))(x) \Leftrightarrow P_{r,t} = P_{r,s} P_{s,t} \end{aligned}$$

In probability theory, the r.h.s. semigroup property is also called the Chapman-Kolmogorov formula. The law of the random states $\eta_t = \text{Law}(X_t)$ satisfy the linear evolution equation

$$\forall s \leq t \qquad \eta_t = \eta_s P_{s,t}$$

Using (5.8), we have the local description of the predictable increment

$$\begin{aligned} [P_{t,t+dt} - Id](f)(x) &= \mathbb{E}\left([f(X_{t+dt}) - f(X_t)] \mid X_t = x\right) \\ &= L_t(f)(x)\ dt \quad \Longleftrightarrow \quad P_{t,t+dt} = Id + L_t\ dt\ + \mathrm{o}(dt) \end{aligned}$$

We quote 2 important consequences of these expansions

$$\frac{d}{dt} P_{s,t} = \frac{1}{dt}\ [\underbrace{P_{s,t+dt}}_{= P_{s,t} P_{t,t+dt}} - P_{s,t}] = P_{s,t} \left[\frac{P_{t,t+dt} - Id}{dt}\right] = P_{s,t} L_t \tag{5.28}$$

$$\frac{d}{ds} P_{s,t} = \frac{1}{-ds}\ [\underbrace{P_{s-ds,t}}_{= P_{s-ds,s} P_{s,t}} - P_{s,t}] = \left[\frac{P_{s-ds,s} - Id}{-ds}\right] = -L_s P_{s,t} \tag{5.29}$$

For time homogeneous models $P_{s,t} = P_{0,t-s} := P_{t-s}$, these two formulae reduce to

$$\frac{d}{dt}P_t = P_t L = LP_t \tag{5.30}$$

The semigroup P_t is sometimes written in the exponential form

$$P_t = e^{tL} \tag{5.31}$$

For pure jump models on a finite set E, the infinitesimal generator L is given by the matrix

$$L(f)(x) = \lambda(x) \sum_{y \in E} (f(y) - f(x))\ S(x,y)$$

where $S(x,y)$ stands for the (x,y)-entry of a stochastic matrix S.

In this situation (5.31) reduces to the exponential matrix formula

$$P_t = e^{tL} = \sum_{n \geq 0} \frac{t^n}{n!}\ L^n \tag{5.32}$$

It is also instructive to notice that

$$\mathbb{P}\left(X_{t+dt} = y \mid X_t = x\right) = P_{dt}(x,y) = e^{Ldt}(x,y)$$

with

$$\begin{array}{rcl} e^{Ldt}(x,y) & \simeq_{dt \downarrow 0} & 1_{x=y} + L(x,y)dt \\ & = & 1_{x=y} + \lambda(x)\ (S(x,y) - 1_{x=y})\ dt = \lambda(x)dt\ S(x,y) + (1 - \lambda(x)dt)\ 1_{x=y} \end{array}$$

This shows that the discrete time approximation of the continuous time jump process is defined by *an aperiodic Bernoulli type jump Markov chain model.*

5.5 Feynman-Kac perturbation semigroups

In the further development of this section X_t stands for the jump-diffusion process with jump intensity λ_t discussed in section 5.2, and $\varphi_{s,t}(x)$ stands for the stochastic flow of the diffusion process (5.1).

5.5.1 Feynman-Kac models

We let $P^{\lambda}_{s,t}$ be the semigroup defined by

$$P^{\lambda}_{s,t}(f)(x) := \mathbb{E}\left[f(\varphi_{s,t}(x))\ \exp\left(-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right)\right] \tag{5.33}$$

Notice that $P^{\lambda}_{s,t}$ is the semigroup of flow of Feynman-Kac measures γ_t defined by

$$\gamma_t(f) = \mathbb{E}\left[f(\varphi_{0,t}(x))\ \exp\left(-\int_0^t \lambda_u(\varphi_{0,u}(x))du\right)\right] = \gamma_s\left(P^{\lambda}_{s,t}(f)\right) \tag{5.34}$$

for any $s \leq t$. The normalized version of these distributions are given by

$$\eta_t(f) = \gamma_t(f)/\gamma_t(1)$$

For the null intensity function $\lambda_t = 0$, $P^0_{s,t}$ coincides with the evolution semigroup of the d-dimensional diffusion without jumps; that is, we have that

$$P^0_{s,t}(f)(x) = \mathbb{E}\left[f(\varphi_{s,t}(x))\right] = P_{s,t}(f)(x)$$

We emphasize that γ_t, and resp. η_t, are the marginal w.r.t. the final time horizon t of the measures $\boldsymbol{\gamma_t}$, and resp. $\boldsymbol{\eta_t}$, on the set of càdlàg paths $E_t = D([0,t],\mathbb{R}^d)$ from the interval $[0,t]$ into $\mathbb{R}^d$ defined for any measurable function $\boldsymbol{f_t}$ on $D([0,t],\mathbb{R}^d)$ by

$$\boldsymbol{\gamma_t(f_t)} = \mathbb{E}\left(\boldsymbol{f_t(X_t)}\ \exp\left(-\int_0^t \boldsymbol{\lambda_s(X_s)}ds\right)\right) \quad \text{and} \quad \boldsymbol{\eta_t(f_t)} = \boldsymbol{\gamma_t(f_t)/\gamma_t(1)}$$

with

$$\boldsymbol{X_t} = (X_s)_{s\leq t} \quad \text{and} \quad \boldsymbol{\lambda_t(X_t)} = \lambda_t(X_t)$$

The measures $\boldsymbol{\eta_t} = \mathbb{Q}_t$ are sometimes written in terms of the weighted distribution

$$\mathbb{Q}_t(dx) := \frac{1}{\mathcal{Z}_t}\ \exp\left(-\int_0^t \lambda_s(x_s)ds\right)\ \mathbb{P}_t(dx) \tag{5.35}$$

of the random trajectories of the Markov chain $\mathbb{P}_t = \text{Law}\,(\boldsymbol{X_t})$, with the normalizing constants

$$\mathcal{Z}_t = \gamma_t(1) = \mathbb{E}\left(\exp\left(-\int_0^t \lambda_s(X_s)ds\right)\right)$$

In the above display, $dx = d(x_s)_{s\leq t}$ stands for an infinitesimal neighborhood of the path $x = (x_s)_{s\leq t}$.

5.5.2 Partition functions

We underline that the normalizing constants $\mathcal{Z}_t$ (also termed in physics the partition function or the free energy), as well as the measures $\mathbb{Q}_t$, can be represented in terms of the flow of marginal measures $(\eta_s)_{0\leq s\leq t}$. More precisely, we have the easily checked multiplicative formulae

$$\mathcal{Z}_t = \mathbb{E}\left(\exp\left(-\int_0^t \lambda_s(X_s)ds\right)\right) = \exp\left(-\int_0^t \eta_s(\lambda_s)ds\right) \tag{5.36}$$

We check this claim using the fact that

$$\frac{d}{dt}\log \mathcal{Z}_t = -\frac{1}{\mathcal{Z}_t}\ \mathbb{E}\left(\lambda_t(X_t)\ \exp\left(-\int_0^t \lambda_s(X_s)ds\right)\right) = \eta_t(\lambda_t)$$

More generally, using this formula the unnormalized measures $\gamma_t := \mathcal{Z}_t \times \eta_t$ can also be rewritten in the following form

$$\gamma_t(f) = \eta_t(f_t)\ \exp\left(-\int_0^t \eta_s(\lambda_s)ds\right) \tag{5.37}$$

Inversely, we also have that

$$\begin{aligned}\eta_t(f_t) &= \gamma_t(f)\ \exp\int_0^t \eta_s(\lambda_s)ds = \mathbb{E}\left(f(X_t)\ \exp\left(-\int_0^t [\lambda_s(X_s)-\eta_s(\lambda_s)]ds\right)\right)\\ &= \mathbb{E}\left(f(X_t)\ \exp\left(-\int_0^t \overline{\lambda}_s(X_s)ds\right)\right)\end{aligned} \tag{5.38}$$

with the normalized function

$$\overline{\lambda}_t = \lambda_t - \eta_t(\lambda_t)$$

5.5.3 Evolution semigroups

Using the fact that

$$\frac{d}{dt}\exp\left(-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right) = -\lambda_t(\varphi_{s,t}(x))\ \exp\left(-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right)$$

we find the integral formula

$$\exp\left(-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right) = 1 - \int_s^t \lambda_r(\varphi_{s,r}(x))\ \exp\left(-\int_s^r \lambda_u(\varphi_{s,u}(x))du\right)\ dr$$

from which we prove that

$$\begin{aligned}P^\lambda_{s,t}(f)(x) &= P^0_{s,t}(f)(x) - \int_s^t \mathbb{E}\left(\mathbb{E}\left(f(\varphi_{s,t}(x))\,|\varphi_{s,r}(x)\right)\lambda_r(\varphi_{s,r}(x))\ \exp\left[-\int_s^r \lambda_u(\varphi_{s,u}(x))du\right]\right)\ dr\\ &= P^0_{s,t}(f)(x) - \int_s^t P^\lambda_{s,r}\left(\lambda_r\ P^0_{r,t}(f)\right)\ dr\end{aligned}$$

Using the above formula, it is readily checked that

$$\begin{aligned}\frac{d}{dt}P^\lambda_{s,t}(f)(x) &= P^0_{s,t}(L_t(f))(x) - \int_s^t P^\lambda_{s,r}\left(\lambda_r\ P^0_{r,t}(L_t(f))\right)\ dr - P^\lambda_{s,t}\left(\lambda_t\ f\ \right)\\ &= P^\lambda_{s,t}(L_t(f) - \lambda_t\ f)(x)\end{aligned}$$

This shows that

$$\frac{d}{dt}P^\lambda_{s,t} = P^\lambda_{s,t}L^\lambda_t \quad \text{with} \quad L^\lambda_t(f) = L_t(f) - \lambda_t\ f \tag{5.39}$$

It is also readily checked that the Feynman-Kac measures γ_t defined in (5.34) satisfy the equation

$$(5.39) \Rightarrow \frac{d}{dt}\gamma_t(f) = \gamma_t(L_t^\lambda(f)) \quad \text{and} \quad \begin{aligned} \tfrac{d}{dt}\eta_t(f) &= \eta_t(L_t^{\overline{\lambda}}(f)) \\ &= \eta_t(L_t(f)) - [\eta_t(\lambda_t f) - \eta_t(f)\eta_t(\lambda_t)] \end{aligned} \tag{5.40}$$

with

$$L_t^{\overline{\lambda}}(f) = L_t(f) - \overline{\lambda}_t\ f$$

For any $y \in \mathbb{R}^d$, we also have that

$$\frac{d}{ds}\exp\left(-\int_s^t \lambda_u(\varphi_{0,u}(y))du\right) = \lambda_s(\varphi_{0,s}(y))\ \exp\left(-\int_s^t \lambda_u(\varphi_{0,u}(y))du\right)$$

from which we prove that

$$1 - \exp\left(-\int_s^t \lambda_u(\varphi_{0,u}(y))du\right) = \int_s^t \lambda_r(\varphi_{0,r}(y))\ \exp\left(-\int_r^t \lambda_u(\varphi_{0,u}(y))du\right)\ dr$$

Recalling that

$$\forall s \leq u \leq t \qquad \varphi_{0,u}(y) = \varphi_{s,u}(\varphi_{0,s}(y))$$

for any $x(= \varphi_{0,s}(y))$ we have

$$\exp\left[-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right] = 1 - \int_s^t \lambda_r(\varphi_{s,r}(x))\ \exp\left[-\int_r^t \lambda_u(\varphi_{s,u}(x))du\right]\ dr$$

This implies that

$$\begin{aligned} &f(\varphi_{s,t}(x))\ \exp\left(-\int_s^t \lambda_u(\varphi_{s,u}(x))du\right) \\ &= f(\varphi_{s,t}(x)) - \int_s^t \lambda_r(\varphi_{s,r}(x))\ f(\varphi_{r,t}(\varphi_{s,r}(x)))\ \exp\left[-\int_r^t \lambda_u(\varphi_{r,u}(\varphi_{s,r}(x)))du\right]\ dr \end{aligned}$$

and therefore

$$P_{s,t}^\lambda(f) = P_{s,t}^0(f) - \int_s^t P_{s,r}^0\left(\lambda_r\ P_{r,t}^\lambda(f)\right)\ dr$$

Using the above formula, it is readily checked that

$$\frac{d}{ds}P_{s,t}^\lambda(f) = -L_s(P_{s,t}^0(f)) + \int_s^t L_s\left(P_{s,r}^0\left(\lambda_r\ P_{r,t}^\lambda(f)\right)\right)\ dr + \lambda_s\ P_{s,t}^\lambda(f)$$

This implies that

$$\frac{d}{ds}P_{s,t}^\lambda = -L_s^\lambda P_{s,t}^\lambda$$

We also have the decomposition

$$P_{s,t}(f)(x) = \mathbb{E}\left(f(X_t)\ 1_{T^{(s)}>t} \mid X_s = x\right) + \mathbb{E}\left(f(X_t)\ 1_{T^{(s)}\leq t} \mid X_s = x\right)$$

with the first time of the jump after time s. By construction, we have

$$\begin{aligned}\mathbb{E}\left(f(X_t)\ 1_{T^{(s)}>t} \mid X_s = x\right) &= \mathbb{E}\left(f(\varphi_{s,t}(x))\ 1_{T^{(s)}>t} \mid X_s = x\right) \\ &= \mathbb{E}\left(f(\varphi_{s,t}(x))\ \exp\left(-\int_s^t \lambda_r(\varphi_{s,r}(x))\ dr\right)\right) = P^{\lambda}_{s,t}(f)(x)\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}\left(f(X_t)\ 1_{T^{(s)}\leq t} \mid X_s = x\right) &= \mathbb{E}\left(\underbrace{\mathbb{E}\left(f(X_t) \mid T^{(s)}, X_{T^{(s)}}\right)}\ 1_{T^{(s)}\leq t} \mid X_s = x\right) \\ &= \mathbb{E}\left(\overbrace{P_{T^{(s)},t}(f)(X_{T^{(s)}})}^{\|}\ 1_{T^{(s)}\leq t} \mid X_s = x\right)\end{aligned}$$

Using (5.6), we prove that

$$\begin{aligned}&\mathbb{E}\left(f(X_t)\ 1_{T^{(s)}\leq t} \mid X_s = x\right) \\ &= \int_s^t \mathbb{E}\left(\left[\int_S S_u(\varphi_{s,u}(x), dy) P_{u,t}(f)(y)\right] \lambda_u(\varphi_{s,u}(x))\ \exp\left[-\int_s^u \lambda_r(\varphi_{s,r}(x)) dr\right] | X_s = x\right)\ du \\ &= \int_s^t P^{\lambda}_{s,u}\left(\lambda_u\ S_u(P_{u,t}(f))\right)(x)\ du\end{aligned}$$

We conclude that

$$P_{s,t}(f)(x) = P^{\lambda}_{s,t}(f)(x) + \int_s^t P^{\lambda}_{s,u}\left(\lambda_u\ S_u(P_{u,t}(f))\right)(x)\ du \tag{5.41}$$

5.5.4 Discrete time approximations

For time homogeneous models $L^{\lambda}_s = L^{\lambda}$ we conclude that

$$P^{\lambda}_{s,t} = P^{\lambda}_{0,t-s} := P^{\lambda}_{t-s} \Rightarrow \frac{d}{dt} P^{\lambda}_t = L^{\lambda} P^{\lambda}_t = P^{\lambda}_t L^{\lambda} \tag{5.42}$$

As in (5.31), the semigroup P^{λ}_t is sometimes written in the exponential form

$$P^{\lambda}_t = e^{tL^{\lambda}}$$

For the pure jump models on a finite set E discussed in (5.32) the Feynman-Kac semigroup (5.42) associated with some potential function V reduces to the exponential matrix formula

$$P_t^V = e^{tL^V} = \sum_{n\geq 0} \frac{t^n}{n!} (L^V)^n \quad \text{with} \quad L^V(x,y) = L(x,y) - V(x)\ 1_{x=y}$$

In this case, we also have

$$\begin{aligned} e^{L^V dt}(x,y) &\simeq_{dt\downarrow 0} 1_{x=y} + L^V(x,y)dt = 1_{x=y} + L(x,y)dt - V(x)\ 1_{x=y}\ dt \\ &\simeq_{dt\downarrow 0} (1 - V(x)dt)\ (1_{x=y} + L(x,y)dt) \\ &\simeq_{dt\downarrow 0} G_{dt}(x)\ P_{dt}(x,y) \quad \text{with} \quad G^{(dt)}(x) := e^{-V(x)dt} \end{aligned}$$

Thus, a discrete time approximation $P_{t_n}^{V,\epsilon}$ of the Feynman-Kac semigroup $P_{t_n}^V = P_{n\epsilon}^V$ on a time mesh $(t_n)_{n\geq 0}$ with $(t_n - t_{n-1}) = \epsilon \simeq 0$ is given by the formulae

$$P_{t_n}^{V,\epsilon}(f)(x) = \mathbb{E}\left(f(X_{t_n}^\epsilon) \prod_{0\leq k<n} G^\epsilon(X_{t_k}^\epsilon) \right)$$

where $X_{t_n}^\epsilon$ stands for a Markov chain with transition probabilities $P_\epsilon(x,y)$

5.6 Integro-differential equations

Continuous time Markov processes are intimately related to intro-differential linear evolution equations. To underline the role of Monte Carlo simulation of discrete generation Markov chain models in the numerical solving of these equations, this section provides a brief description of the continuous time version of the Markov transport Equation (4.14).

Using elementary manipulations, we can show that the evolution equation of the laws η_t of the random states X_t of the jump diffusion model (5.3) is given by weak integro-differential equation

$$\frac{d}{dt}\eta_t(f) = \eta_t\left(L_t(f)\right) \tag{5.43}$$

for sufficiently regular functions f on $E = \mathbb{R}^d$, with the integro-differential operator L_t defined in (5.15).

We further assume that $S_t(x,dy) = q_t(x,y)\ dy$ and the law of the random states $\eta_t(dy) = p_t(y)\ dy$ have a smooth density $q_t(x,y)$, and $p_t(y)$ w.r.t. the Lebesgue measure dy on $\mathbb{R}^d$. In this situation, the equation (5.43) takes the form

$$\frac{d}{dt}\eta_t(f) = \int f(x)\ \frac{dp_t}{dt}(x)\ dx = \int p_t(x)\ L_t^c(f)(x)\ dx + \int p_t(x)\ \lambda_t(x)\ q_t(x,y)\ (f(y) - f(x))\ dxdy$$

For smooth functions f with compact support, we have the integration by part formulae

$$\int p_t(x)\ L_t^c(f)(x)\ dx = \int L_t^{c,\star}(p_t)(x)\ f(x)\ dx$$

with the dual differential operator

$$L_t^{c,\star}(p_t) = -\sum_{i=1}^d \partial_{x_i}\left(a'_{t,i}\ p_t\right) + \frac{1}{2}\sum_{i,j=1}^d \partial_{x_i,x_j}\left(\left(\sigma'_t(\sigma'_t)^T\right)_{i,j}\ p_t\right)$$

On the other hand, we also have that

$$\int p_t(x)\ \lambda_t(x)\ q_t(x,y)\ (f(y) - f(x))\ dxdy = \int f(x)\ L_t^{d,\star}(p_t)(x)$$

with

$$L_t^{d,\star}(p_t)(x) := \left(\int p_t(y)\ \lambda_t(y)\ q_t(y,x)\ dy \right) - p_t(x)\ \lambda_t(x)$$

This implies that

$$\int f(x)\ \frac{dp_t}{dt}(x)\ dx = \int f(x)\ L_t^\star(p_t)(x)\ dx \quad \text{with} \quad L_t^\star = L_t^{c,\star} + L_t^{d,\star}$$

Since this equation is valid for any smooth functions f, we conclude that

$$\frac{dp_t}{dt}(x) = L_t^\star(p_t)(x) \tag{5.44}$$

Inversely, any equation of the form (5.44) can be interpreted as the probability densities of the random states of a jump diffusion model of the form (5.3).

The Fokker-Planck type integro-differential equation (5.44) is sometimes rewritten in the following form

$$\frac{dp_t}{dt}(x) + \text{div}\,(a_t\ p_t) - \frac{1}{2}\nabla^2 : \left(\sigma_t(\sigma_t)^T p_t\right) - \Theta_t(p_t) = 0$$

with the operators

$$\begin{aligned}
\text{div}\,(a_t\ p_t) &:= \sum_{i=1}^d \partial_i\,(a_{t,i}\ p_t) \\
\nabla^2 : \left(\sigma_t(\sigma_t)^T p_t\right) &= \sum_{i,j=1}^d \partial_{i,j}\left(\left(\sigma_t(\sigma_t)^T\right)_{i,j}\ p_t\right) \\
\Theta_t(p_t) &:= \int p_t(y)\ \lambda_t(y)\ [q_t(y,x)\ dy - \delta_x(dy)]
\end{aligned}$$

5.7 Stability properties

5.7.1 Invariant measures

A probability measure π on some state space E is invariant w.r.t. some time homogeneous Markov semigroup $P_t = P_{0,t}$ if we have $\pi P_t = \pi$, for any $t \geq 0$. The invariance property is also characterized in terms of the generator L of the semigroup:

$$\pi \text{ is } P_t\text{-invariant} \iff (\forall t \geq 0 \quad \pi P_t = \pi) \iff \pi L = 0$$

We check this claim using the couple of formulae

$$\pi\left(\frac{P_t - Id}{t}\right) \to_{t \downarrow 0} \pi L \quad \text{and} \quad P_t = Id + \int_0^t L P_s\ ds \tag{5.30}$$

The measure π is reversible w.r.t. P_t if we have

$$\pi(dx)\ P_t(x,dy) = \pi(dy)\ P_t(y,dx)$$

or equivalently, for any pair of functions $(f,g) \in \mathcal{B}(S)^2$

$$\forall t \geq 0 \quad \pi(f P_t(g)) = \pi(P_t(f) g) \iff \pi(f L(g)) = \pi(L(f) g)$$

5.7.2 Foster-Lyapunov conditions

We start with a general result that allows to apply the V-norm contraction techniques developed in section 4.4.4.

We consider the sg $P_t = P_{0,t}$ of a time homogeneous Markov process on some state space E satisfying the evolution equations (5.30) for some infinitesimal generator L acting on some domain $D(L)$ of sufficiently smooth functions.

We assume that there exists some non negative function $W \in D(L)$ on E such that

$$L(W) \leq -a\ W\ +\ c \tag{5.45}$$

for some parameters $a > 0$, and $c \geq 0$. In this situation, for any $t > 0$ the Markov transition P_t satisfies the Foster-Lyapunov condition (4.40)

$$P_t(W) \leq \epsilon_t\ W + c_t \quad \text{with} \quad \epsilon_t = \frac{1}{1+at} \quad \text{with} \quad c_t = c\ t \tag{5.46}$$

In addition, if P_t satisfies the Dobrushin local contraction condition (4.39) for any $t > 0$ then there exists an unique invariant measure $\pi = \pi M$.
Furthermore, for any $h > 0$ there exist some positive function V s.t. $\beta_V(P_h) < 1$. In particular, we have the exponential contraction inequality

$$\|P_{h\lfloor t/h\rfloor}(x,.) - \pi\|_V \leq \beta_V(P_h)^{\lfloor t/h\rfloor}\ (1 + V(x) + \pi(V))$$

To prove these claims, we first use (5.28) to check that

$$\begin{aligned} P_t(W) &= W + \int_0^t P_s(L(W))\ ds \\ &\leq W + \int_0^t [-a\ P_s(W) + c]\ ds = W + ct - a\int_0^t P_s(W)ds \end{aligned}$$

On the other hand, by an integration by part we have

$$\begin{aligned} \int_0^t P_s(W)ds &= [s\ P_s(W)]_0^t - \int_0^t s\ \frac{d}{ds}P_s(W)\ ds \\ &= t\ P_t(W) - \int_0^t s\ P_s(\underbrace{L(W)}_{\leq c})\ ds \geq t\ P_t(W) - ct^2/2 \end{aligned}$$

This implies that

$$P_t(W) \leq W + ct - a\left(t\ P_t(W) - ct^2/2\right)$$

from which we conclude that

$$P_t(W) \leq \frac{1}{1+at}\ W + ct\ \frac{1+at/2}{1+at} \leq \frac{1}{1+at}\ W + ct$$

The last assertion is a direct consequence of the theorem 4.51 applied to the Markov transition $P_{nh} = P_h^n$. This ends the proof of the desired result.

Replacing in (5.45) W by $W + b$ for some $b > 0$ we find that

$$L(W+b) = L(W) + b\ \underbrace{L(1)}_{=0} = L(W) \leq -a\ (W+b)\ +\ (c+ab)$$

This shows that there is no loss of generality to assume that (5.45) is met for some function $W \geq b$, for some $b > 0$.

On the other hand, using the fact that

$$(5.45) \quad \text{with} \quad c > 0 \quad \Rightarrow \quad L\left(\frac{W}{c}\right) \leq -a\,\frac{W}{c} \; + \; 1$$

there is also no loss of generality to assume that (5.45) is met for some non negative function $W \geq 0$, with $c = 1$.

Last, but not least it suffices to check that there exists some subset $A \subset E$ s.t.

$$\forall x \in E - A \quad W^{-1}(x)L(W)(x) \leq -a \quad \text{and} \quad \forall x \in A \quad L(W)(x) \leq c$$

In this situation, it is readily checked that

$$L(W) = W^{-1}L(W)1_{E-A}\; W \; + \; L(W)\; 1_A \Rightarrow L(W) \leq -a\; W \; + c \tag{5.47}$$

For instance, when E is equipped with some norm $\|.\|$, and $L(W)$ is continuous it suffices to find some sufficiently large radius R s.t.

$$\forall \|x\| \geq R \qquad W^{-1}(x)L(W)(x) \leq -a \tag{5.48}$$

We also mention that any s.g. P_t on $E = \mathbb{R}^d$ satisfies the Dobrushin local contraction condition (4.39) for any $t > 0$ as soon as the Markov transitions $P_t(x, dy) = p_t(x, y)\; dy$ have continuous densities $(x, y) \mapsto p_t(x, y) > 0$ w.r.t. the Lebesgue measure dy. Much more is true, rephrasing (4.42) we have the following result.

We assume that the s.g. P_t satisfies some minorization property

$$P_t(x, dy) \geq q_t(x, y)\; dy \tag{5.49}$$

for some function $q_t(x, y)$ that is lower semicontinuous w.r.t. the first variable, and upper semicontinuous w.r.t. the second. In this situation the s.g. P_t satisfies the Dobrushin local contraction condition (4.39) for any $t > 0$.

We consider a d-dimensional jump diffusion process with jump intensity function $\lambda(x)$, jump amplitude transition $M(x, dy)$ and a stochastic flow $\varphi_{s,t}(x)$. We assume that the s.g. P_t^0 of the stochastic flow satisfies the minorisation condition (5.49) for some $q_t^0(x, y)$. When the intensity function λ is bounded the s.g. P_t of the jump-diffusion model satisfies (5.49) with

$$P_t(x, dy) \geq e^{-\|\lambda\| t}\; q_t^0(x, y)\; dy$$

The proof if a direct consequence of the integral formula (5.41):

$$P_t(f) \quad = \quad P_t^\lambda(f) + \int_0^t P_u^\lambda(\lambda S P_{t-u}(f))\; du \geq P_t^\lambda(f) \geq e^{-\|\lambda\| t}\; P_t^0(f)$$

for any non negative function f. This clearly ends the proof of the desired result. ∎

5.8 Some illustrations

It is clearly far beyond the scope of these lecture notes to discuss in full details the absolutely continuity properties of Markov semigroups. For a more detailed discussion we refer the reader to [150, 330, 372], and references therein.

In this section we present some examples of Lyapunov functions for some classes of diffusion processes.

5.8.1 Ornstein-Uhlenbeck process

The infinitesimal generator of the Ornstein-Uhlenbeck process is given by

$$L := -\sum_{i=1}^{d} a_i \; x^i \; \partial_i + \frac{1}{2} \sum_{i,j=1}^{d} \left(\sigma(\sigma)^T\right)_j^i \; \partial_{i,j}$$

with some deterministic covariance matrix σ and a collection of parameters $a_i < \infty$. In this situation, the s.g. $P_t(x, dy) = p_t(x, y)\, dy$ has smooth densities $(x, y) \mapsto p_t(x, y) > 0$ w.r.t. the Lebesgue measure dy [330]. When $a_{min} = \wedge_{1 \leq i \leq d} a_i > 0$, we can choose the Lyapunov function

$$W(x) = \frac{1}{2} \; \|x\|^2 := \frac{1}{2} \sum_{1 \leq i \leq d} (x^i)^2$$

In this situation, we have

$$\begin{aligned} L(W)(x) &= -\sum_{i=1}^{d} a_i \; (x^i)^2 + \frac{1}{2} \; \text{Trace}\left(\sigma(\sigma)^T\right) \\ &\leq -\; 2 \; a_{min} \; W(x) + \text{Trace}\left(\sigma\sigma^T\right) \end{aligned}$$

5.8.2 Stochastic gradient process

The infinitesimal generator of the stochastic gradient process is given by

$$L := -\sum_{i=1}^{d} \partial_i V \; \partial_i + \frac{1}{2} \sum_{i=1}^{d} \partial_{i,i}$$

with some smooth function V behaving as $\|x\|^\alpha$ with $\alpha \geq 1$ at infinity; that is, there exists some sufficiently large radius R s.t. for any $\|x\| \geq R$ we have

$$\sum_{i=1}^{d} \; |\partial_{i,i} V(x)| \leq c_1 \; \|x\|^{\alpha - 2} \quad \text{and} \quad \sum_{i=1}^{d} \; (\partial_i V(x))^2 \geq c_2 \; \|x\|^{2(\alpha-1)}$$

for some constants $c_1 < \infty$ and $c_2 > 0$. In this situation, the s.g. $P_t(x, dy) = p_t(x, y)\, dy$ has smooth densities $(x, y) \mapsto p_t(x, y) > 0$ w.r.t. the Lebesgue measure dy [330]. In addition, we can choose the Lyapunov functions

$$W(x) = \exp\left(2\epsilon V\right)$$

for any $\epsilon \in]0, 1[$. To check this claim, we observe that

$$\begin{aligned} & W^{-1} L(W) \\ & = -2\epsilon \textstyle\sum_{i=1}^{d} \; (\partial_i V)^2 + \frac{1}{2} \; \sum_{i=1}^{d} \; \left(4\epsilon^2 \, (\partial_i V)^2 + 2\epsilon \partial_{i,i} V\right) \\ & \leq -2\epsilon \left[(1 - \epsilon) \, \|\nabla V\|^2 - \tfrac{1}{2} \; \triangle V\right] \end{aligned}$$

Under our assumptions, for any $\|x\| \geq R > [c_1/(2c_2)]^{1/\alpha}$ we have

$$\begin{aligned}(1-\epsilon)\,\|\nabla V\|^2 - \frac{1}{2}\,\triangle V &\geq (1-\epsilon)c_2\,\|x\|^{2(\alpha-1)} - \frac{c_1}{2}\,\|x\|^{\alpha-2} \\ &= \|x\|^{\alpha-2}\,\left[c_2\,\|x\|^{\alpha} - \frac{c_1}{2}\right] \\ &\geq R^{\alpha-2}\,\left[c_2\,R^{\alpha} - \frac{c_1}{2}\right] > 0\end{aligned}$$

We conclude that (5.48) holds with $a := 2\epsilon R^{\alpha-2}\,\left[c_2\,R^{\alpha} - \frac{c_1}{2}\right]$.

5.8.3 A two dimensional Langevin diffusion

We consider the $\mathbb{R}^2$-valued stochastic process $X_t = (q_t, p_t)$ defined by the couple of equations

$$\left\{\begin{array}{rcl} dq_t &=& \beta\,\dfrac{p_t}{m}\,dt \\ dp_t &=& -\beta\,\left(\dfrac{\partial V}{\partial q}(q_t) + \dfrac{\sigma^2}{2}\,\dfrac{p_t}{m}\right)\,dt + \sigma\,dW_t \end{array}\right. \tag{5.50}$$

with some positive constants β, m, σ, some Brownian motion W_t, and some smooth positive function V on $\mathbb{R}$ s.t. for some sufficiently large R we have

$$|q| \geq R \qquad q\frac{\partial V}{\partial q}(q) \geq \delta\,\left(V(q) + q^2\right)$$

for some positive constant δ. This condition is clearly met when V behaves at infinity as $q^{2\alpha}$, for some $\alpha \geq 1$; that is, there exists some sufficiently large radius R s.t. for any $|q| \geq R$ we have

$$q\frac{\partial V}{\partial q}(q) \geq c_1\,q^{2\alpha} \quad \text{and} \quad c_2\,q^{2\alpha} \geq V(q)$$

for some constants $c_1 < \infty$ and $c_2 > 0$.

The generator of the process (5.50) is defined by

$$L(f)(q,p) = \beta\,\frac{p}{m}\,\frac{\partial f}{\partial q} - \beta\,\left(\frac{\partial V}{\partial q} + \frac{\sigma^2}{2}\,\frac{p}{m}\right)\frac{\partial f}{\partial p} + \frac{\sigma^2}{2}\,\frac{\partial^2 f}{\partial p^2}$$

We let $W(q,p)$ be the function on $\mathbb{R}^2$ defined by

$$W(q,p) = \frac{1}{2m}\,p^2 + V(q) + \frac{\epsilon}{2}\,\left(\frac{\sigma^2}{2}\,q^2 + 2pq\right) \quad \text{with} \quad \epsilon < \frac{\sigma^2}{2m}$$

Recalling that $2pq \leq p^2 + q^2$, we prove that

$$\begin{aligned} W(q,p) &\leq \frac{1}{2}\left(\frac{1}{m} + \epsilon\right)\,p^2 + \frac{\epsilon}{2}\,\left(\frac{\sigma^2}{2} + 1\right)\,q^2 + V(q) \\ &\leq C^\star(\epsilon)\,\left(p^2 + q^2 + V(q)\right) \end{aligned}$$

with

$$C^\star(\epsilon) := \max\left\{\frac{1}{2}\left(\frac{1}{m} + \epsilon\right), \frac{\epsilon}{2}\,\left(\frac{\sigma^2}{2} + 1\right), 1\right\}$$

On the other hand, we have

$$\begin{aligned} L(W) &= \beta\frac{p}{m}\left(\frac{\partial V}{\partial q} + \epsilon\,\frac{\sigma^2}{2}\,q + \epsilon\,p\right) \\ &\qquad -\beta\,\left(\frac{\partial V}{\partial q} + \frac{\sigma^2}{2}\,\frac{p}{m}\right)\,\left(\frac{p}{m} + \epsilon\,q\right) + \frac{\sigma^2}{2m} \\ &= -\beta\,\left[\frac{1}{m}\,\left(\frac{\sigma^2}{2m} - \epsilon\right)\,p^2 + \epsilon\,q\,\frac{\partial V}{\partial q}\right] + \frac{\sigma^2}{2m} \end{aligned}$$

Under our assumptions, this implies that for any $|q| \geq R$ we have

$$\begin{aligned} L(W) &\leq -\beta \left[\frac{1}{m}\left(\frac{\sigma^2}{2m} - \epsilon\right) p^2 + \epsilon\ \delta\ \left(V(q) + q^2\right)\right] + \frac{\sigma^2}{2m} \\ &\leq -C_\star(\epsilon,\delta)\ \left(p^2 + q^2 + V(q)\right) + \frac{\sigma^2}{2m} \end{aligned}$$

with

$$C_\star(\epsilon,\delta) := \beta\ \min\left\{\left(\frac{1}{m}\left(\frac{\sigma^2}{2m} - \epsilon\right), \epsilon\ \delta\right)\right\}$$

We conclude that for any $|q| \geq R$ we have

$$\begin{aligned} (W^{-1}L(W))(q,p) &\leq -\frac{C_\star(\epsilon,\delta)\ \left(p^2 + q^2 + V(q)\right) - \frac{\sigma^2}{2m}}{W(q,p)} \\ &\leq -\frac{C_\star(\epsilon,\delta)\ \left(p^2 + q^2 + V(q)\right) - \frac{\sigma^2}{2m}}{C^\star(\epsilon)\ (p^2 + q^2 + V(q))} \\ &= -\frac{C_\star(\epsilon,\delta)}{C^\star(\epsilon)} + \frac{\sigma^2}{2mC^\star(\epsilon)}\ \frac{1}{p^2 + q^2 + V(q)} \\ &\leq -\left[\frac{C_\star(\epsilon,\delta)}{C^\star(\epsilon)} - \frac{\sigma^2}{2mC^\star(\epsilon)}\ \frac{1}{p^2 + q^2}\right] \\ &\leq -\left[\frac{C_\star(\epsilon,\delta)}{C^\star(\epsilon)} - \frac{\sigma^2}{2mC^\star(\epsilon)}\ \frac{1}{p^2 + R^2}\right] \end{aligned}$$

Choosing $\overline{R}$ sufficiently large s.t.

$$|p| \geq \overline{R} \quad \text{and} \quad |q| \geq R$$

$$\Rightarrow \tfrac{C_\star(\epsilon,\delta)}{C^\star(\epsilon)} - \tfrac{\sigma^2}{2mC^\star(\epsilon)}\ \tfrac{1}{p^2+q^2} \geq \tfrac{C_\star(\epsilon,\delta)}{C^\star(\epsilon)} - \tfrac{\sigma^2}{2mC^\star(\epsilon)}\ \tfrac{1}{\overline{R}^2+R^2} \geq \tfrac{C_\star(\epsilon,\delta)}{2C^\star(\epsilon)} > 0$$

we conclude that (5.47) is met with the set

$$A = \{|p| \wedge |q| < \overline{R} \vee R\}$$

and the parameter $a = C_\star(\epsilon,\delta)/(2C^\star(\epsilon)) > 0$; that is, we have that

$$|p| \wedge |q| \geq \overline{R} \vee R \Rightarrow (W^{-1}L(W))(q,p) \leq -a$$

Chapter 6

Diffusions on manifolds

6.1 A review of differential geometry

6.1.1 Projection operators

We let $\mathcal{V} = \text{Vect}\,(V_1, \ldots, V_p) \subset \mathbb{R}^r$ be a p-dimensional vector space with a (non necessarily orthonormal) basis $(V_1, \ldots, V_p) \in (\mathbb{R}^r)^p$, with the column vectors $V_i = \begin{bmatrix} V_i^1 \\ \vdots \\ V_i^r \end{bmatrix}$, with $1 \leq i \leq p \leq r$. We equip $\mathcal{V}$ with the Euclidian scalar product

$$g_{i,j} := \langle V_i, V_j \rangle = V_i^T V_j = \sum_{1 \leq k \leq r} V_i^k V_k^j = \text{tr}(V_i V_j^T)$$

We let $g^{i,j}$ be the entries of the inverse g^{-1} of the matrix $g = (g_{i,j})_{1 \leq i,j \leq p}$ and we set

$$V = [V_1, \ldots, V_p] \quad \text{and} \quad V^T := \begin{bmatrix} V_1^T \\ \vdots \\ V_p^T \end{bmatrix} \implies g = V^T V$$

In this notation, the orthogonal projection $\pi_{\mathcal{V}} : W \in \mathbb{R}^r \mapsto \pi_{\mathcal{V}}(W) \in \mathcal{V}$ on $\mathcal{V}$ is given by the matrix

$$\pi_{\mathcal{V}} = V g^{-1} V^T \implies \pi_{\mathcal{V}}(W) = \sum_{1 \leq i \leq p} \left\langle \sum_{1 \leq j \leq p} g^{i,j} V_j, W \right\rangle V_i$$

The r.h.s. formula comes from the fact that

$$g^{-1} V^T = \begin{bmatrix} \sum_{1 \leq j \leq p} g^{1,j} V_j^T \\ \vdots \\ \sum_{1 \leq j \leq p} g^{p,j} V_j^T \end{bmatrix} \implies g^{-1} V^T W = \begin{bmatrix} \sum_{1 \leq j \leq p} g^{1,j} V_j^T W \\ \vdots \\ \sum_{1 \leq j \leq p} g^{p,j} V_j^T W \end{bmatrix}$$

Given a collection of vectors $(W_i)_{1 \leq i \leq k}$ we set

$$\pi_{\mathcal{V}}([W_1, \ldots, W_k]) = [\pi_{\mathcal{V}}(W_1), \ldots, \pi_{\mathcal{V}}(W_k)]$$

In this notation, for any $W_1, W_2 \in \mathbb{R}^r$, we observe that

$$W_1 W_2^T := \left[W_2^1\ W_1, \dots W_2^r\ W_1\right] \Longrightarrow \pi_{\mathcal{V}}(W_1 W_2^T) = \left[W_2^1\ \pi_{\mathcal{V}}(W_1), \dots W_2^r\ \pi_{\mathcal{V}}(W_1)\right] = \pi_{\mathcal{V}}(W_1) W_2^T$$

In summary, we have proved that

$$\pi_{\mathcal{V}}(W_1 W_2^T) = \pi_{\mathcal{V}}(W_1) W_2^T$$

If we choose an orthonormal basis $(U_1, \dots, U_p) \in (\mathbb{R}^r)^p$ we have

$$V_i = \sum_{1 \le j \le p} \langle V_i, U_j \rangle U_j = \underbrace{[U_1, \dots, U_p]}_{U} \underbrace{\begin{bmatrix} U_1^T \\ \vdots \\ U_p^T \end{bmatrix}}_{U^T} V_i := U U^T V_i \implies V = UP$$

with

$$P = U^T V = \begin{bmatrix} \langle U_1, V_1 \rangle & \dots & \langle U_1, V_p \rangle \\ \vdots & & \\ \langle U_p, V_1 \rangle & \dots & \langle U_p, V_p \rangle \end{bmatrix}$$

from which we conclude that

$$\begin{aligned} V g^{-1} V^T &= V(V^T V) - 1 V^T \\ &= UP((UP)^T UP)^{-1}(UP)^T = UP(P^T U^T UP)^{-1}(UP)^T \\ &= UP(P^T P)^{-1} P^T U^T = UPP^{-1}(P^T)^{-1} P^T U^T = UU^T \end{aligned}$$

This shows that the projection matrix $\pi_{\mathcal{V}} = \left(\pi^k_{\mathcal{V},l}\right)_{1 \le k,l \le r}$ doesn't depend on the choice of the basis of the vector field $\mathcal{V}$, and we have that

$$\pi_{\mathcal{V}} = \pi_{\mathcal{V}}^T \quad \text{and} \quad \pi_{\mathcal{V}} \pi_{\mathcal{V}} = \pi_{\mathcal{V}} \Rightarrow \forall 1 \le k, l \le r \quad \sum_{1 \le i \le r} \pi^i_{\mathcal{V},l} \pi^i_{\mathcal{V},k} = \pi^k_{\mathcal{V},l}$$

6.1.2 First order covariant derivatives

We further assume that we are given a collection of smooth vector functionals (a.k.a. vector fields) $V_i : x \in \mathbb{R}^r \mapsto V_i(x) \in \mathbb{R}^r$ and $V_j^\perp : x \in \mathbb{R}^r \mapsto V_j^\perp(x) \in \mathbb{R}^r$, with $1 \le i \le p$ and $1 \le j \le q = r - p$ such that

$$\mathbb{R}^{p+q} = \underbrace{\mathrm{Vect}\,(V_1, \dots, V_p)}_{=\mathcal{V}} \overset{\perp}{+} \underbrace{\mathrm{Vect}\left(V_1^\perp, \dots, V_q^\perp\right)}_{=\mathcal{V}^\perp} \quad \left(\Leftrightarrow \langle V_i, V_j^\perp \rangle = 0 \quad \forall\ 1 \le i \le p \quad \text{and} \quad \forall\ 1 \le j \le q\right)$$

By construction, for any vector field we have

$$W = \pi_{\mathcal{V}}(W) + \pi_{\mathcal{V}^\perp}(W) \tag{6.1}$$

Notice that $\pi := \pi_{\mathcal{V}}$ and $\pi_\perp := \pi_{\mathcal{V}^\perp}$ are smooth matrix functionals and

$$\pi_\perp(W) = \sum_{1 \le i \le q} \left\langle \sum_{1 \le j \le q} g_\perp^{i,j} V_j^\perp, W \right\rangle V_i^\perp \tag{6.2}$$

with the entries $g_\perp^{i,j}$ of the inverse $g_\perp^{-1}$ of the matrix $g_\perp = (g_{\perp,i,j})_{1\leq i,j\leq p}$ given by

$$g_{\perp,i,j} := \langle V_i^\perp, V_j^\perp \rangle$$

Given a smooth function F and a smooth vector field $W = \begin{bmatrix} W^1 \\ \vdots \\ W^r \end{bmatrix}$ on $\mathbb{R}^r$, for any $1 \leq i \leq r$

we set

$$\partial F = \begin{bmatrix} \partial_{x_1} F \\ \vdots \\ \partial_{x_r} F \end{bmatrix} \qquad \partial_{x_i} W = \begin{bmatrix} \partial_{x_i} W^1 \\ \vdots \\ \partial_{x_i} W^r \end{bmatrix} \quad \text{and} \quad \partial W = \left[\partial W^1, \ldots, \partial W^r\right]$$

The Euclidian gradient operator $\partial \ : \ F \mapsto \partial F$ maps smooth functions to vector fields $\partial F \ : \ x \mapsto (\partial F)(x)$ that encapsulate information about the change of the function F w.r.t. infinitesimal variations of the independent coordinates x_i of the state $x = (x_1, \ldots, x_r)^T$.

We also consider the operators

$$\partial_W(F) = \sum_{1\leq k\leq r} W^k \ \partial_{x_k}(F) = W^T \partial F = \langle W, \partial F \rangle \quad \text{and} \quad \nabla = \pi \partial$$

Given smooth curve $C \ : \ t \in [0,1] \mapsto C(t) = \left(C^1(t), \ldots, C^r(t)\right)^T \in \mathbb{R}^r$ starting at some state $C(0) = x \in \mathbb{R}^r$, with a velocity vector field W, we have

$$\frac{d}{dt} F(C(t)) = \sum_{1\leq k\leq r} W^k(C(t)) \ \partial_{x_k}(F)(C(t)) = \left(\partial_W(F)\right)(C(t)) = \langle W(C(t)), (\partial F)(C(t)) \rangle$$

The function $\partial_W(F)$ is called the directional derivative of F w.r.t. the vector field W. The r.h.s. equation makes clear the dependency of the gradient on the inner product structure on $\mathbb{R}^r$. If $A((\partial F)(x), W(x))$ represents the angle between $(\partial F)(x)$ and $W(x)$ we have

$$\partial_W(F)(x) = \|(\partial F)(x)\| \ \ \|W(x)\| \ \ \cos\left(A((\partial F)(x), W(x))\right)$$

When $W(x)$ is perpendicular to $(\partial F)(x)$, then the rate of change of $(\partial F)(x)$ in the direction $W(x)$ is null. In the reverse angle, the rate of change of $(\partial F)(x)$ in the direction $W(x)$ is maximal when $W(x)$ is parallel to $(\partial F)(x)$.

The covariant derivative

$$\nabla F := \pi\left(\partial F\right) = \partial F - \pi_\perp(\partial F)$$

expresses the changes of the function F w.r.t. vectors $W \in \mathcal{V}$:

$$\forall W \in \mathcal{V} \qquad \langle \partial F, W \rangle = \langle \pi\left(\partial F\right), W \rangle = \langle \nabla F, W \rangle$$

By construction, we readily check that

$$\partial(FW) = \partial F \ W^T + F \ \partial W \tag{6.3}$$

and

$$\begin{aligned}\nabla(FW) &= \pi\partial(FW)\\ &= \pi\left(\partial F\ W^T\right)+\pi\left(F\ \partial W\right)\\ &= \pi\left(\partial F\right)\ W^T+F\ \pi\left(\partial W\right)\end{aligned}$$

from which we conclude that

$$\boxed{\nabla(FW)=\nabla F\ W^T+F\ \nabla W} \tag{6.4}$$

We also observe that

$$\begin{aligned}(6.1) &\implies \partial W=\partial\pi(W)+\partial\pi_\perp(W)\\ &\implies \nabla W=\nabla\pi(W)+\nabla\pi_\perp(W)\end{aligned} \tag{6.5}$$

On the other hand, we have

$$\pi(W)=\begin{bmatrix}\pi_1^1 & \dots & \pi_r^1\\ \vdots & \vdots & \vdots\\ \pi_1^r & \dots & \pi_r^r\end{bmatrix}\begin{bmatrix}W^1\\ \vdots\\ W^r\end{bmatrix}=\begin{bmatrix}\sum_{1\le k\le r}\pi_k^1\ W^k\\ \vdots\\ \sum_{1\le k\le r}\pi_k^r\ W^k\end{bmatrix}$$

and therefore

$$\begin{aligned}\partial\ \pi(W) &=\\ &= \left[\sum_{1\le k\le r}\partial\left(\pi_k^1\ W^k\right),\dots,\sum_{1\le k\le r}\partial\left(\pi_k^r\ W^k\right)\right]\\ &= \left[\sum_{1\le k\le r}W^k\ \partial\pi_k^1,\dots,\sum_{1\le k\le r}W^k\ \partial\pi_k^r\right]+\left[\sum_{1\le k\le r}\pi_k^1\ \partial W^k,\dots,\sum_{1\le k\le r}\pi_k^r\ \partial W^k\right]\\ &= \sum_{1\le k\le r}W^k\ \left[\partial\pi_k^1,\dots,\partial\pi_k^r\right]+\begin{bmatrix}\partial_{x_1}W^1 & \dots & \partial_{x_1}W^r\\ \vdots & \vdots & \vdots\\ \partial_{x_r}W^1 & \dots & \partial_{x_r}W^r\end{bmatrix}\begin{bmatrix}\pi_1^1 & \dots & \pi_r^1\\ \vdots & \vdots & \vdots\\ \pi_r^1 & \dots & \pi_r^r\end{bmatrix}\\ &= \sum_{1\le k\le r}W^k\ \partial\pi_k+\left[\partial W^1,\dots,\partial W^r\right]\pi^T=\sum_{1\le k\le r}W^k\ \partial\pi_k+(\partial W)\pi^T\end{aligned}$$

with the vector field

$$\pi_k=\begin{bmatrix}\pi_k^1\\ \vdots\\ \pi_k^r\end{bmatrix}$$

This yields the formula

$$\boxed{\nabla\pi(W)\ =\ \pi\partial\pi(W)=\sum_{1\le k\le r}W^k\ \nabla\pi_k+\pi\partial W\pi^T}$$

with the matrix

$$\pi\partial W\pi^T = \begin{bmatrix} \sum_{1\leq l\leq r}\pi_l^1\sum_{1\leq k\leq r}\pi_k^1\,\partial_{x_l}W^k, & \dots & ,\sum_{1\leq l\leq r}\pi_l^1\sum_{1\leq k\leq r}\pi_k^r\,\partial_{x_l}W^k \\ \vdots & \vdots & \vdots \\ \sum_{1\leq l\leq r}\pi_l^r\sum_{1\leq k\leq r}\pi_k^1\,\partial_{x_l}W^k, & \dots & ,\sum_{1\leq l\leq r}\pi_l^r\sum_{1\leq k\leq r}\pi_k^r\,\partial_{x_l}W^k \end{bmatrix}$$

On the other hand, we have

$$\begin{aligned} \mathrm{tr}\left(\pi\partial W\pi^T\right) &= \sum_{1\leq k,l\leq r}\left(\sum_{1\leq i\leq r}\pi_l^i\pi_k^i\right)\,\partial_{x_l}W^k \\ &= \sum_{1\leq k,l\leq r}\pi_k^l\,\partial_{x_l}W^k = \mathrm{tr}\left(\pi\partial W\right) = \mathrm{tr}\left(\nabla W\right) = \sum_{1\leq k\leq r}\partial_{\pi_k}W^k \end{aligned} \tag{6.6}$$

with

$$\begin{aligned} \nabla W &= \pi\partial W = \left[\pi(\partial W^1),\dots,\pi(\partial W^r)\right] = \left[\nabla W^1,\dots,\nabla W^r\right] \\ &= \pi^T\partial W = \begin{bmatrix}\pi_1^T\\ \vdots\\ \pi_r^T\end{bmatrix}\left[\partial W^1,\dots,\partial W^r\right] = \begin{pmatrix} \partial_{\pi_1}W^1, & \dots & ,\partial_{\pi_1}W^r \\ \vdots & \vdots & \vdots \\ \partial_{\pi_r}W^1, & \dots & ,\partial_{\pi_r}W^r \end{pmatrix} \end{aligned}$$

In much the same way, we have

$$\begin{aligned} \nabla\pi_k &= \pi\partial\pi_k = \pi^T\partial\pi_k \\ &= \begin{bmatrix}\pi_1^T\\ \vdots\\ \pi_r^T\end{bmatrix}\left[\partial\pi_k^1,\dots,\partial\pi_k^r\right] = \begin{bmatrix} \partial_{\pi_1}\pi_k^1, & \dots, & \partial_{\pi_1}\pi_k^r \\ \vdots & \vdots & \vdots \\ \partial_{\pi_r}\pi_k^1, & \dots, & \partial_{\pi_r}\pi_k^r \end{bmatrix} \end{aligned}$$

We conclude that

$$\begin{aligned} \mathrm{tr}\left(\nabla\pi(W)\right) &= \sum_{1\leq k\leq r}W^k\,\mathrm{tr}\left(\nabla\pi_k\right) + \sum_{1\leq k\leq r}\partial_{\pi_k}W^k \\ &= \sum_{1\leq k,l\leq r}W^k\,\partial_{\pi_l}\pi_k^l + \mathrm{tr}\left(\pi\partial W\right) = \sum_{1\leq k,l\leq r}W^k\,\partial_{\pi_l}\pi_k^l + \mathrm{tr}\left(\nabla W\right) \end{aligned}$$

Using the fact that

$$\begin{aligned} (6.5)\Longrightarrow\ \mathrm{tr}\left(\nabla W\right) &= \mathrm{tr}\left(\nabla\pi(W)\right) + \mathrm{tr}\left(\nabla\pi_\perp(W)\right) \\ &= \textstyle\sum_{1\leq k,l\leq r}W^k\,\partial_{\pi_l}\pi_k^l + \mathrm{tr}\left(\nabla W\right) + \mathrm{tr}\left(\nabla\pi_\perp(W)\right) \end{aligned}$$

we conclude that

$$\mathrm{tr}\left(\nabla\pi_\perp(W)\right) = -\sum_{1\leq k,l\leq r}W^k\,\partial_{\pi_l}\pi_k^l$$

6.1.3 Divergence and mean curvature

For $q=1$, we have $g_\perp = \left\|V_1^\perp\right\|^2$, $g_\perp^{-1} = \left\|V_1^\perp\right\|^{-2}$ and

$$\pi_\perp(W) = \left\langle\overline{V}_1^\perp, W\right\rangle\,\overline{V}_1^\perp = \overline{V}_1^\perp\,\overline{V}_1^{\perp,T}W = \frac{V_1^\perp\,V_1^{\perp,T}}{V_1^{\perp,T}V_1^\perp}\,W \quad\text{with}\quad \overline{V}_1^\perp = \frac{V_1^\perp}{\left\|V_1^\perp\right\|}$$

In this particular case, we have the formula

$$\mathrm{tr}\left(\nabla\pi_\perp(W)\right) = \langle \mathbb{H}, W\rangle \tag{6.7}$$

with the mean curvature vector $\mathbb{H}$ defined by

$$\mathbb{H} = \mathrm{div}_\perp\left(\overline{V}_1^\perp\right)\ \overline{V}_1^\perp = -\sum_{1\leq k\leq r}\left(\sum_{1\leq l\leq r}\partial_{\pi_l}\pi_k^l\right)\ e_k \quad \text{with} \quad \mathrm{div}_\perp\left(\overline{V}_1^\perp\right) = \sum_{1\leq i\leq r}\partial_{x_i}\left(\overline{V}_1^{\perp,i}\right)$$

and the unit vectors e_i on $\subset \mathbb{R}^r$ defined by

$$\forall 1\leq i\leq r \qquad e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-th term}$$

To check this assertion, we use the fact that

$$\partial\pi_\perp(W) \ = \ \partial\left(\left\langle \overline{V}_1^\perp, W\right\rangle\right)\ \overline{V}_1^{\perp,T} + \left\langle \overline{V}_1^\perp, W\right\rangle\ \partial\overline{V}_1^\perp$$

from which we prove that

$$\nabla\pi_\perp(W) \ = \ \pi\partial\pi_\perp(W) = \pi\left(\partial\left\langle \overline{V}_1^\perp, W\right\rangle\right)\ \overline{V}_1^{\perp,T} + \left\langle \overline{V}_1^\perp, W\right\rangle\ \pi\left(\partial\overline{V}_1^\perp\right)$$

To analyze the r.h.s. term, we observe that

$$\pi\left(\partial\overline{V}_1^\perp\right) = \left[\partial\overline{V}_1^{\perp,1},\ldots,\partial\overline{V}_1^{\perp,r}\right] - \left[\pi_\perp\left(\partial\overline{V}_1^{\perp,1}\right),\ldots,\pi_\perp\left(\partial\overline{V}_1^{\perp,r}\right)\right]$$

On the other hand, we have that

$$\mathrm{tr}\left(\pi\left(\partial\left\langle \overline{V}_1^\perp, W\right\rangle\right)\ \overline{V}_1^{\perp,T}\right) = \left\langle \pi\left(\partial\left\langle \overline{V}_1^\perp, W\right\rangle\right), \overline{V}_1^\perp\right\rangle = 0$$

and

$$\mathrm{tr}\left(\pi\left(\partial\overline{V}_1^\perp\right)\right) = \sum_{1\leq i\leq r}\partial_{x_i}\overline{V}_1^{\perp,i} - \mathrm{tr}\left[\pi_\perp\left(\partial\overline{V}_1^{\perp,1}\right),\ldots,\pi_\perp\left(\partial\overline{V}_1^{\perp,r}\right)\right]$$

Finally, we check that

$$\begin{aligned}
\mathrm{tr}\left[\pi_\perp\left(\partial\overline{V}_1^{\perp,1}\right),\ldots,\pi_\perp\left(\partial\overline{V}_1^{\perp,r}\right)\right] &= \mathrm{tr}\left[\left\langle \overline{V}_1^\perp, \partial\overline{V}_1^{\perp,1}\right\rangle\ \overline{V}_1^\perp,\ldots,\left\langle \overline{V}_1^\perp, \partial\overline{V}_1^{\perp,r}\right\rangle\ \overline{V}_1^\perp\right] \\
&= \sum_{1\leq i\leq r}\left\langle \overline{V}_1^\perp, \partial\overline{V}_1^{\perp,i}\right\rangle\ \overline{V}_1^{\perp,i} \\
&= \sum_{1\leq j\leq r}\overline{V}_1^{\perp,j}\sum_{1\leq i\leq r}\partial_{x_j}\overline{V}_1^{\perp,i}\ \overline{V}_1^{\perp,i} \\
&= \sum_{1\leq j\leq r}\overline{V}_1^{\perp,j}\left\langle \partial_{x_j}\overline{V}_1^\perp, \overline{V}_1^\perp\right\rangle = \frac{1}{2}\sum_{1\leq j\leq r}\overline{V}_1^{\perp,j}\partial_{x_j}\left\langle \overline{V}_1^\perp, \overline{V}_1^\perp\right\rangle = 0
\end{aligned}$$

This ends the proof of (6.7).

 Our next objective is to extend this formula to any dimensional vector spaces $\mathcal{V}^\perp$ spanned by a given basis of vector fields $V_i^\perp$, $1 \leq i \leq q$. In this general situation, we have

$$\pi_\perp(W) = \sum_{1\leq i\leq q} \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle \; V_i^\perp$$

$$\Rightarrow \; \partial\pi_\perp(W) = \sum_{1\leq i\leq q} \left[\partial \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle\right] \; V_i^{\perp,T} + \sum_{1\leq i\leq q} \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle \; \partial V_i^\perp$$

$$\Rightarrow \; \nabla\pi_\perp(W) = \sum_{1\leq i\leq q} \pi\left[\partial \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle\right] \; V_i^{\perp,T} + \sum_{1\leq i\leq q} \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle \; \nabla V_i^\perp$$

Using the fact that

$$\mathrm{tr}\left(\pi\left[\partial\left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W\right\rangle\right] \; V_i^{\perp,T}\right) = \left\langle \pi\left[\partial\left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W\right\rangle\right], V_i^\perp \right\rangle = 0$$

we prove the formula

$$\begin{aligned} \mathrm{tr}\left(\nabla\pi_\perp(W)\right) &= \sum_{1\leq i\leq q} \left\langle \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp, W \right\rangle \; \mathrm{tr}\left(\nabla V_i^\perp\right) \\ &= \left\langle \sum_{1\leq j\leq q} \left[\sum_{1\leq i\leq q} g_\perp^{i,j} \; \mathrm{tr}\left(\nabla V_i^\perp\right)\right] V_j^\perp, W \right\rangle \end{aligned} \tag{6.8}$$

We also observe that

$$\begin{aligned} \nabla V_i^\perp &= \pi\partial V_i^\perp = \partial V_i^\perp - \pi_\perp \partial V_i^\perp \\ &= \left[\partial V_i^{\perp,1}, \ldots, \partial V_i^{\perp,r}\right] - \left[\pi_\perp \partial V_i^{\perp,1}, \ldots, \pi_\perp \partial V_i^{\perp,r}\right] \end{aligned}$$

with

$$\pi_\perp \partial V_i^{\perp,j} = \sum_{1\leq k\leq q} \left\langle \sum_{1\leq l\leq q} g_\perp^{k,l} V_l^\perp, \partial V_i^{\perp,j} \right\rangle \; V_k^\perp$$

This yields

$$\mathrm{tr}\left(\nabla V_i^\perp\right) = \sum_{1\leq m\leq r} \partial_{x_m} V_i^{\perp,m} - \sum_{1\leq m\leq r} \sum_{1\leq k,l\leq q} g_\perp^{k,l} \left\langle V_l^\perp, \partial V_i^{\perp,m} \right\rangle \; V_k^{\perp,m}$$

and therefore

$$\sum_{1\leq i\leq q} g_\perp^{i,j} \; \mathrm{tr}\left(\nabla V_i^\perp\right) = \sum_{1\leq i\leq q} g_\perp^{i,j} \left[\sum_{1\leq m\leq r} \partial_{x_m} V_i^{\perp,m} - \sum_{1\leq m\leq r} \sum_{1\leq k,l\leq q} g_\perp^{k,l} \left\langle V_l^\perp, \partial V_i^{\perp,m} \right\rangle \; V_k^{\perp,m}\right]$$

We further assume that the vector fields $V_i^\perp$ satisfy the following condition

$$\forall 1 \leq i \leq q \quad \forall 1 \leq k,l \leq r \qquad \partial_{x_k} V_i^{\perp,l} = \partial_{x_l} V_i^{\perp,k}$$

This condition is clearly met when for gradient type vector field models $V_i^\perp = \partial\varphi_i$ associated with some smooth functions φ_i on $\mathbb{R}^r$. In this situation, we have

$$\begin{aligned}\left\langle \partial_{x_i} V_k^\perp, V_l^\perp \right\rangle &= \sum_{1\leq j\leq r} \partial_{x_i} V_k^{\perp,j} V_l^{\perp,j} \\ &= \sum_{1\leq j\leq r} V_l^{\perp,j}\ \partial_{x_j} V_k^{\perp,i} = \partial_{V_l^\perp} V_k^{\perp,i} = \left\langle\ \partial V_k^{\perp,i}, V_l^\perp \right\rangle\end{aligned}$$

and

$$\begin{aligned}\left\langle \partial_{V_j^\perp} V_k^\perp, V_l^\perp \right\rangle &= \sum_{1\leq i\leq r} V_j^i \left\langle \partial_{x_i} V_k^\perp, V_l^\perp \right\rangle \\ &= \sum_{1\leq i,i'\leq r} V_j^i\ \partial_{x_i}\ V_k^{\perp,i'}\ V_l^{\perp,i'} = \sum_{1\leq i,i'\leq r} V_j^i\ \partial_{x_i'}\ V_k^{\perp,i}\ V_l^{\perp,i'} \\ &= \sum_{1\leq i,i'\leq r} V_l^{\perp,i'}\ \partial_{x_i'}\ V_k^{\perp,i}\ \ V_j^i\ = \left\langle \partial_{V_l^\perp} V_k^\perp, V_j^\perp \right\rangle \end{aligned} \tag{6.9}$$

Using the fact that

$$\partial_{x_j}\left(g_{\perp,k,m}\right) = \partial_{x_j}\left\langle V_k^\perp, V_m^\perp\right\rangle = \left\langle \partial_{x_j} V_k^\perp, V_m^\perp\right\rangle + \left\langle V_k^\perp, \partial_{x_j} V_m^\perp\right\rangle$$

we have

$$\begin{aligned}\partial_{V_j^\perp}\left(g_{\perp,k,m}\right) &= \left\langle \partial_{V_j^\perp} V_k^\perp, V_m^\perp\right\rangle + \left\langle \partial_{V_j^\perp} V_m^\perp, V_k^\perp\right\rangle \\ &= \left\langle \partial_{V_m^\perp} V_k^\perp, V_j^\perp\right\rangle + \left\langle \partial_{V_k^\perp} V_m^\perp, V_j^\perp\right\rangle = \left\langle \partial_{V_m^\perp} V_k^\perp + \partial_{V_k^\perp} V_m^\perp, V_j^\perp\right\rangle\end{aligned} \tag{6.10}$$

In addition, for any $1 \leq k, l, m \leq q$ we have

$$\left\langle \partial_{V_m^\perp} V_k^\perp, V_l^\perp\right\rangle = \sum_{1\leq i\leq r} V_m^{\perp,i}\left\langle \partial_{x_i} V_k^\perp, V_l^\perp\right\rangle = \sum_{1\leq i\leq r} V_m^{\perp,i}\left\langle\ \partial V_k^{\perp,i}, V_l^\perp\right\rangle \tag{6.11}$$

This yields

$$\begin{aligned}\sum_{1\leq i\leq q} g_\perp^{i,j}\ \mathrm{tr}\left(\nabla V_i^\perp\right) &= \sum_{1\leq i\leq q}\sum_{1\leq m\leq r} g_\perp^{i,j}\ \partial_{x_m} V_i^{\perp,m} - \sum_{1\leq i,k,l\leq q} g_\perp^{i,j}\ g_\perp^{k,l} \sum_{1\leq m\leq r} V_k^{\perp,m}\ \left\langle \partial V_i^{\perp,m}, V_l^\perp\right\rangle \\ &= \sum_{1\leq i\leq q}\sum_{1\leq m\leq r} g_\perp^{i,j}\left[\ \partial_{x_m} V_i^{\perp,m} - \sum_{1\leq k,l\leq q} g_\perp^{k,l}\ \left\langle \partial_{V_k^\perp} V_i^\perp, V_l^\perp\right\rangle\right]\end{aligned} \tag{6.12}$$

To get one step further in our discussion, we need to recall some basic facts on the differentiation of determinant of invertible matrices. We let $\epsilon \mapsto A(\epsilon) = \begin{bmatrix} A_1^1(\epsilon) & \ldots & A_r^1(\epsilon) \\ \vdots & \vdots & \vdots \\ A_1^r(\epsilon) & \ldots & A_r^r(\epsilon)\end{bmatrix}$ be a smooth $(r\times r)$-invertible matrix functional. The co-factor expansion of the determinant of $A(\epsilon)$ along the i-th row is given by the formula

$$\det(A(\epsilon)) = \sum_{1\leq j\leq r} A_j^i(\epsilon)\ C_j^i(\epsilon) \Longrightarrow C_j^i(\epsilon) = \frac{\partial \det(A)}{\partial A_j^i}(\epsilon)$$

where $C_j^i(\epsilon)$ stands for the co-factor of the entry $A_j^i(\epsilon)$ defined by multiplying by $(-1)^{i+j}$ the determinant of the minor of the entry in the i-th row and the j-th column. We recall that this (i,j)-minor is the determinant of the sub-matrix deduced from $A(\epsilon)$ by deleting the i-th row and j-th column.

The inverse of the matrix $A(\epsilon)$ is defined by

$$A^{-1}(\epsilon) = \frac{1}{\det(A(\epsilon))} \begin{bmatrix} C_1^1(\epsilon) & \dots & C_1^r(\epsilon) \\ \vdots & \vdots & \vdots \\ C_r^1(\epsilon) & \dots & C_r^r(\epsilon) \end{bmatrix} = \frac{1}{\det(A(\epsilon))}\ C^T(\epsilon)$$

This leads quickly to the Jacobi formula for the derivative of the detreminant

$$\begin{aligned} \frac{d}{d\epsilon}\left(\det(A(\epsilon))\right) &= \sum_{1\leq i,j\leq r} \frac{\partial \det(A)}{\partial A_j^i}(\epsilon)\ \frac{dA_j^i(\epsilon)}{d\epsilon} = \sum_{1\leq i\leq r}\sum_{1\leq j\leq r} (C^T(\epsilon))_i^j\ \left(\frac{dA(\epsilon)}{d\epsilon}\right)_j^i \\ &= \mathrm{tr}\left(C^T(\epsilon)\ \frac{dA(\epsilon)}{d\epsilon}\right) \\ &= \det(A(\epsilon))\ \mathrm{tr}\left(A^{-1}(\epsilon)\ \frac{dA(\epsilon)}{d\epsilon}\right) \end{aligned}$$

For any smooth vector field W on $\mathbb{R}^r$ we set

$$\mathrm{div}_\perp(W) = \frac{1}{\sqrt{\det(g_\perp)}} \sum_{1\leq m\leq r} \partial_{x_m}\left(\sqrt{\det(g_\perp)}\ W^m\right) \tag{6.13}$$

We have

$$\begin{aligned} \frac{1}{\sqrt{\det(g_\perp)}}\ \partial_{x_m}\left(\sqrt{\det(g_\perp)}\ W^m\right) &= \frac{1}{\sqrt{\det(g_\perp)}}\partial_{x_m}\left(\sqrt{\det(g_\perp)}\right)\ W^m + \partial_{x_m}(W^m) \\ &= \frac{1}{2\det(g_\perp)}\partial_{x_m}\left(\det(g_\perp)\right)\ W^m + \partial_{x_m}(W^m) \\ &= \frac{1}{2}\ \mathrm{tr}\left(g_\perp^{-1}\partial_{x_m} g_\perp\right) W^m + \partial_{x_m}(W^m) \\ &= \frac{1}{2}\sum_{1\leq k,l\leq q} g_\perp^{k,l}\partial_{x_m} g_{\perp,k,l}\ W^m + \partial_{x_m}(W^m) \end{aligned}$$

from which we find that

$$\frac{1}{\sqrt{\det(g_\perp)}}\ \partial_{x_m}\left(\sqrt{\det(g_\perp)}\ W^m\right) = \sum_{1\leq k,l\leq q} g_\perp^{k,l}\left\langle \partial_{x_m} V_k^\perp, V_l^\perp\right\rangle\ W^m + \partial_{x_m}(W^m)$$

On the other hand, we have

$$\begin{aligned} \sum_{1\leq j\leq q} g_{\perp,i,j}\ g_\perp^{j,k} = 1_{i=k} \ &\Rightarrow\ \sum_{1\leq j\leq q} g_{\perp,i,j}\ \partial_{x_m} g_\perp^{j,k} = -\sum_{1\leq j\leq q}\left(\partial_{x_m} g_{\perp,i,j}\right)\ g_\perp^{j,k} \\ &\Rightarrow\ \sum_{1\leq i,j\leq q} g_\perp^{l,i} g_{\perp,i,j}\ \partial_{x_m} g_\perp^{j,k} = \partial_{x_m} g_\perp^{l,k} = -\sum_{1\leq i,j\leq q} g_\perp^{l,i} g_\perp^{k,j}\ \partial_{x_m} g_{\perp,i,j} \end{aligned}$$

Applying these formulas to $W = \sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp$ we find that

$$\begin{aligned}
&\mathrm{div}_\perp\left(\textstyle\sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp\right)\\
&= \sum_{1\leq j,k,l\leq q} g_\perp^{i,j}\, g_\perp^{k,l} \sum_{1\leq m\leq r} V_j^{\perp m}\left\langle \partial_{x_m} V_k^\perp, V_l^\perp\right\rangle\\
&\qquad - \sum_{1\leq j\leq q}\sum_{1\leq m\leq r} V_j^{\perp,m} \sum_{1\leq k,l\leq q} g_\perp^{i,k} g_\perp^{j,l}\, \partial_{x_m} g_{\perp,k,l} + \sum_{1\leq j\leq q} g_\perp^{i,j} \sum_{1\leq m\leq r} \partial_{x_m} V_j^{\perp,m}
\end{aligned}$$

Using (6.9) we obtain

$$\begin{aligned}
&\mathrm{div}_\perp\left(\textstyle\sum_{1\leq j\leq q} g_\perp^{i,j} V_j^\perp\right)\\
&= \sum_{1\leq j\leq q} g_\perp^{i,j} \left\langle \sum_{1\leq k,l\leq q} g_\perp^{k,l}\, \partial_{V_l^\perp} V_k^\perp, V_j^\perp\right\rangle - \sum_{1\leq j\leq q} g_\perp^{i,j} \sum_{1\leq k,l\leq q} g_\perp^{k,l}\, \partial_{V_k^\perp} g_{\perp,j,l} + \sum_{1\leq j\leq q} g_\perp^{i,j} \sum_{1\leq m\leq r} \partial_{x_m} V_j^{\perp,m}
\end{aligned}$$

Using (6.10) we have

$$\sum_{1\leq k,l\leq q} g_\perp^{k,l}\, \partial_{V_k^\perp}(g_{\perp,j,l}) = \sum_{1\leq k,l\leq q} g_\perp^{k,l} \left\langle \partial_{V_k^\perp} V_j^\perp, V_l^\perp\right\rangle + \left\langle \sum_{1\leq k,l\leq q} g_\perp^{k,l} \partial_{V_k^\perp} V_l^\perp, V_j^\perp, \right\rangle$$

Combining this formula with (6.12) we conclude that

$$\begin{aligned}
&\mathrm{div}_\perp\left(\textstyle\sum_{1\leq j\leq q} g_\perp^{i,j}\, V_j^\perp\right)\\
&= \sum_{1\leq j\leq q} g_\perp^{i,j}\left[\sum_{1\leq m\leq r} \partial_{x_m} V_j^{\perp,m} - \sum_{1\leq k,l\leq q} g_\perp^{k,l}\left\langle \partial_{V_k^\perp} V_j^\perp, V_l^\perp\right\rangle\right] = \sum_{1\leq j\leq q} g_\perp^{i,j}\, \mathrm{tr}\left(\nabla V_j^\perp\right)
\end{aligned} \tag{6.14}$$

Finally, using (6.8) we conclude that

$$\mathrm{tr}\left(\nabla \pi_\perp(W)\right) = \mathrm{tr}\left((\pi \partial \pi_\perp) W\right) = \langle \mathbb{H}, W\rangle \tag{6.15}$$

with the mean curvature vector

$$\mathbb{H} = \sum_{1\leq i\leq q} \mathrm{div}_\perp\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\, V_j^\perp\right) V_i^\perp = -\sum_{1\leq k\leq r}\left(\sum_{1\leq l\leq r} \partial_{\pi_l} \pi_k^l\right) e_k \tag{6.16}$$

We let $\mathrm{div}(W)$ be the divergence of a vector field W defined by

$$\mathrm{div}(W) = \mathrm{tr}\left(\nabla W\right)$$

Using (6.5), we have

$$\mathrm{div}(W) = \mathrm{tr}\left(\nabla \pi(W)\right) + \mathrm{tr}\left(\nabla \pi_\perp(W)\right)$$

Choosing $W = \sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp$, for some $1 \leq i \leq q$, we have

$$\mathrm{div}\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right) = \mathrm{tr}\left(\nabla\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right)\right) = \mathrm{div}_\perp\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right) = \sum_{1\leq j\leq q} g_\perp^{i,j}\ \mathrm{div}\left(V_j^\perp\right)$$

We check this claim using the fact that

$$\begin{aligned}\nabla\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right) &= \sum_{1\leq j\leq q} \nabla\left(g_\perp^{i,j}\ V_j^\perp\right)\\ &= \sum_{1\leq j\leq q} \nabla\left(g_\perp^{i,j}\right)\ V_j^{\perp,T} + \sum_{1\leq j\leq q} g_\perp^{i,j}\ \nabla V_j^\perp\end{aligned}$$

so that

$$\mathrm{div}\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right) = \sum_{1\leq j\leq q}\left\langle\nabla\left(g_\perp^{i,j}\right), V_j^\perp\right\rangle + \sum_{1\leq j\leq q} g_\perp^{i,j}\ \mathrm{tr}\left(\nabla V_j^\perp\right) = \sum_{1\leq j\leq q} g_\perp^{i,j}\ \mathrm{tr}\left(\nabla V_j^\perp\right)$$

6.1.4 Laplacian and second order covariant derivatives

By (6.6) we have

$$\mathrm{div}(W) = \mathrm{tr}\left(\nabla W\right) = \sum_{1\leq k\leq r}\partial_{\pi_k} W^k = \mathrm{tr}\left(\nabla\pi(W)\right) + \left\langle\sum_{1\leq i\leq q}\ \mathrm{div}_\perp\left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ V_j^\perp\right)\ V_i^\perp, W\right\rangle$$

If we choose $W = \partial F$ we find the second covariant derivative

$$\nabla^2 F = \sum_{1\leq k\leq r}\partial_{x_k}(F)\ \nabla\pi_k + \pi\partial^2 F\pi^T \tag{6.17}$$

and the Laplacian formula

$$\Delta F := \mathrm{tr}\left(\nabla^2 F\right) = \mathrm{tr}\left(\pi\partial^2 F\right) + \sum_{1\leq k,l\leq r}\partial_{x_k} F\ \partial_{\pi_l}\pi_k^l \tag{6.18}$$

with $\nabla^2 F = \nabla(\nabla F)$, and the Hessian matrix $\partial^2 F = (\partial_{x_k,x_l} F)_{1\leq k,l\leq r}$. On the other hand, we also have

$$\nabla F = \pi\partial F = \pi^T\partial F = \begin{bmatrix}\pi_1^T\\ \vdots\\ \pi_r^T\end{bmatrix}\partial F = \begin{bmatrix}\partial_{\pi_1} F\\ \vdots\\ \partial_{\pi_r} F\end{bmatrix}$$

and

$$\nabla^2 F = \nabla(\nabla F) = [\nabla\partial_{\pi_1} F, \ldots, \nabla\partial_{\pi_r} F] = \begin{pmatrix}\partial_{\pi_1}\partial_{\pi_1} F & \ldots & \partial_{\pi_1}\partial_{\pi_r} F\\ \vdots & \vdots & \vdots\\ \partial_{\pi_r}\partial_{\pi_1} F & \ldots & \partial_{\pi_r}\partial_{\pi_r} F\end{pmatrix}$$

This shows that

$$\mathrm{tr}\left(\nabla^2 F\right) = \sum_{1\leq i\leq r}\partial_{\pi_i}^2 F = \mathrm{tr}\left(\pi\partial^2 F\right) + \sum_{1\leq k,l\leq r}\partial_{x_k} F\ \partial_{\pi_l}\pi_k^l$$

Using (6.15), we also have

$$\Delta F = \mathrm{tr}\left(\nabla^2 F\right) = \sum_{1\leq i\leq r} \partial^2_{\pi_i} F = \mathrm{tr}\left(\pi\partial^2 F\right) - \langle \mathbb{H}, \partial F\rangle = \mathrm{tr}\left(\pi\partial^2 F\right) - \partial_{\mathbb{H}} F \tag{6.19}$$

with the mean curvature vector $\mathbb{H} \in \mathcal{V}^\perp$ defined in (6.16)

6.2 Stochastic differential calculus on manifolds

6.2.1 Embedded manifolds

In this section we briefly recall some terminology used in geometry and differential calculus. We only consider submanifolds S of dimension p which are "smooth subsets" of the ambient Euclidian space $\mathbb{R}^{p+q}$, for some $q \geq 1$. The case $q = 1$ corresponds to hypersurfaces (a.k.a. hypermanifolds). We let $\varphi :\ x \in \mathbb{R}^{r=p+q} \mapsto \varphi(x) = (\varphi_1(x), \ldots, \varphi_q(x))^T \in \mathbb{R}^q$ be a smooth function with a non empty connected null-level set $S := \varphi^{-1}(0)$ s.t.

$$\forall x \in S \qquad \mathrm{rank}\left(\partial\varphi(x)\right) = \left(\partial\varphi_1(x), \ldots, \partial\varphi_q(x)\right) = q$$

We consider a smooth curve $C\ :\ t \in [0,1] \mapsto C(t) = \left(C^1(t), \ldots, C^r(t)\right)^T \in S$ starting at some state $C(0) = x \in S$, with a velocity vector field W, that is we have that

$$\frac{dC}{dt} = \left(\frac{dC^1}{dt}, \ldots, \frac{dC^r}{dt}\right)^T = W(C(t)) = \left(W^1(C(t)), \ldots, W^r(C(t))\right)^T$$

By construction, we have

$$\forall 1 \leq i \leq q \qquad \frac{d}{dt}\varphi_i(C(t)) = \sum_{1\leq j\leq r} \left(\partial_{x_j}\varphi_i\right)(C(t))\ W^j(C(t)) = 0$$

For $t = 0$, this implies that

$$\langle \partial\varphi_i(x), W(x)\rangle = \sum_{1\leq j\leq r} \left(\partial_{x_j}\varphi_i\right)(x)\ W^j(x) = 0 \quad \Longleftrightarrow \quad W(x) \in \ker\left(\partial\varphi_i(x)\right)$$

We let $T_x(S)$ be the vector space spanned by the kernels $\ker\left(\partial\varphi_i(x)\right)$ of the gradient vectors $\partial\varphi_i(x)$, with $1 \leq i \leq q$; that is

$$T_x(S) = \mathrm{Vect}\left(\cup_{1\leq i\leq q}\ker\left(\partial\varphi_i(x)\right)\right)$$

Under our assumptions, we have

$$\mathbb{R}^r = T_x(S) \overset{\perp}{+} T_x^\perp(S) \quad \text{with} \quad T_x^\perp(S) = \mathrm{Vect}\left(\partial\varphi_1(x), \ldots, \partial\varphi_q(x)\right)$$

This implies that $T_x(S)$ is a p-dimensional vector space, so that S is a p-dimensional manifold embedded in the ambient space $\mathbb{R}^r$.

We let $\pi(x)$ be the orthogonal projection from $\mathbb{R}^r$ into $T_x(S)$, and by $\mathbb{H}(x)$ the mean curvature vector given by

$$\mathbb{H} = \sum_{1\leq i\leq q} \mathrm{div}_\perp \left(\sum_{1\leq j\leq q} g_\perp^{i,j}\ \partial\varphi_j \right)\ \partial\varphi_i$$

with

$$g_\perp^{-1} = \left(g_\perp^{i,j}\right)_{1\leq i,j\leq q} \quad \text{and} \quad g_\perp = (g_{\perp,i,j})_{1\leq i,j\leq q} = (\langle \partial\varphi_i, \partial\varphi_j\rangle)_{1\leq i,j\leq q}$$

where $\mathrm{div}_\perp(.)$ stands for the divergence operator defined in (6.13). We also recall from (6.16) that

$$\forall 1\leq k\leq r \qquad \mathbb{H}^k = -\sum_{1\leq i,j\leq r} \pi_j^i \partial_{x_i}\pi_k^j = -\sum_{1\leq j\leq r} \partial_{\pi_j}\pi_j^k \ \Leftrightarrow\ \mathbb{H}^T = -\sum_{1\leq j\leq r} \partial_{\pi_j}\pi_j$$

with the vector fields π_j on $\mathbb{R}^r$ defined by the column vectors

$$\forall 1\leq j\leq r \qquad \pi_j := \begin{bmatrix} \pi_j^1 \\ \vdots \\ \pi_j^r \end{bmatrix}$$

The r.h.s. formulation comes from the fact that $\pi = \pi^T$. By construction, we also notice that

$$\forall 1\leq i\leq q \qquad \partial\varphi_i \in T^\perp(S) \quad \Longrightarrow \quad \nabla\varphi_i = \pi(\partial\varphi_i) = 0 \tag{6.20}$$

In the case of orthogonal constraints

$$\langle \partial\varphi_i, \partial\varphi_j\rangle = 1_{i=j}\ \|\partial\varphi_j\|^2 \Rightarrow g_\perp^{i,j} = 1_{i=j}\ \|\partial\varphi_j\|^{-2} \Rightarrow \mathbb{H} = \sum_{1\leq i\leq q} \mathrm{div}_\perp\left(\frac{\partial\varphi_i}{\|\partial\varphi_i\|^2}\right)\ \partial\varphi_i$$

In addition, by (6.13) we find the computationally useful formula

$$q=1 \Rightarrow\ \mathbb{H} \ = \ \mathrm{div}_\perp\left(\frac{\partial\varphi_1}{\|\partial\varphi_1\|^2}\right)\ \partial\varphi_1 = \left[\sum_{1\leq m\leq r} \partial_{x_m}\left(\frac{\partial_{x_m}\varphi_1}{\|\partial\varphi_1\|}\right)\right]\ \frac{\partial\varphi_1}{\|\partial\varphi_1\|} \tag{6.21}$$

In the special case of the sphere $S = \mathbb{S}^p \subset \mathbb{R}^{p+1}$, with $p = r-1$ we can can take

$$\varphi(x) = \|x\| - 1 \Rightarrow \partial\varphi(x) = x/\|x\| \quad \text{and} \quad \pi(x) = Id - \partial\varphi(x)\partial\varphi(x)^T = Id - \frac{xx^T}{x^Tx} \tag{6.22}$$

In this situation, we have

$$\forall x\neq 0 \qquad \mathbb{H}(x) = \left[\sum_{1\leq m\leq r} \partial_{x_m}\left(\frac{x_m}{\sqrt{x_1^2+\ldots+x_r^2}}\right)\right]\ \frac{x}{\sqrt{x_1^2+x_2^2}} = (r-1)\ \frac{x}{\|x\|^2} = (r-1)\ \frac{x}{x^Tx} \tag{6.23}$$

We check this claim using the fact that

$$\partial_{x_m}\left(\frac{x_m}{\sqrt{x_1^2+\ldots+x_r^2}}\right) = \frac{1}{\sqrt{x_1^2+\ldots+x_r^2}}\left[1 - \frac{x_m^2}{(x_1^2+\ldots+x_r^2)}\right]$$

$$\Rightarrow \mathrm{div}_\perp\left(\frac{\partial\varphi}{\|\partial\varphi\|^2}\right) = \frac{1}{\|x\|}\ \sum_{1\leq m\leq r} \partial_{x_m}\left(\frac{x_m}{\|x\|}\right) = \frac{1}{\|x\|^2}\ \sum_{1\leq m\leq r}\left(1 - \frac{x_m^2}{\|x\|^2}\right) = \frac{r-1}{\|x\|^2}$$

For $p = 2$, the projection on the unit sphere can also be represented in terms of the cross product

$$\pi(x)W(x) = \frac{x}{\|x\|} \wedge W_x \overset{x\in\mathbb{S}^2}{=} x \wedge W_x = \begin{pmatrix} x_2W^3(x) - x_3W^2(x) \\ W^1(x)x_2 - W^2(x)x_1 \\ x_1W^2(x) - x_2W^1(x) \end{pmatrix} \tag{6.24}$$

In much the same way the cylinder of unit radius on $\mathbb{R}^2$ is given $\varphi(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2} - 1 = 0$. In this case, we have

$$\partial\varphi(x) = \frac{1}{\sqrt{x_1^2 + x_2^2}} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \quad \text{and} \quad \pi(x) = Id - \partial\varphi(x)\partial\varphi(x)^T = \begin{pmatrix} \frac{x_2^2}{x_1^2+x_2^2} & -\frac{x_1x_2}{x_1^2+x_2^2} & 0 \\ -\frac{x_1x_2}{x_1^2+x_2^2} & \frac{x_1^2}{x_1^2+x_2^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.25}$$

and

$$\mathbb{H}(x) = \left(\sum_{m=1,2} \partial_{x_m}(\partial_{x_m}\varphi)(x) \right) \partial\varphi(x) = \frac{1}{x_1^2 + x_2^2} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$$

6.2.2 Brownian motion on manifolds

6.2.3 A diffusion model in the ambient space

We let $X_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^r \end{pmatrix}$ be the $\mathbb{R}^r$-valued diffusion defined by

$$dX_t = -\frac{1}{2}\, \mathbb{H}(X_t)\, dt + \pi(X_t)\, dB_t \tag{6.26}$$

where B_t stands for a standard r-dimensional Brownian motion.

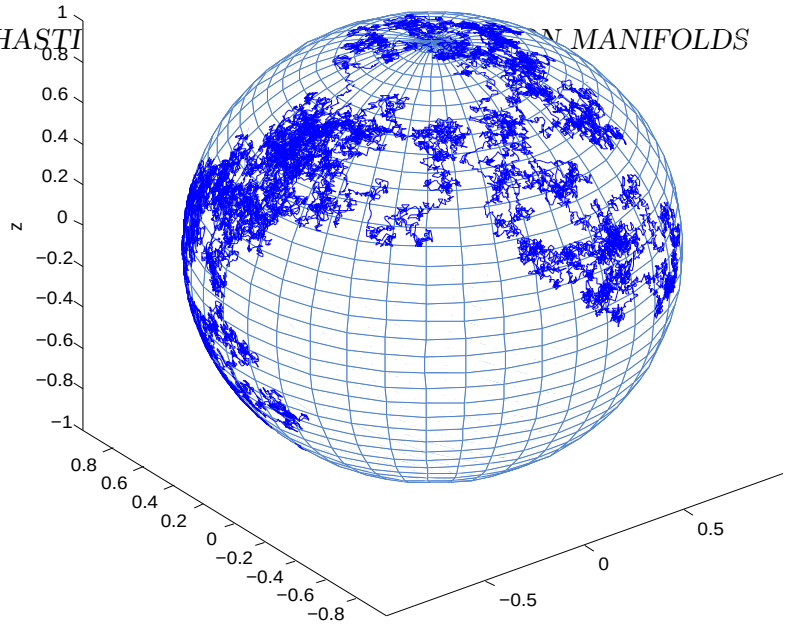
In the special case of the sphere $S = \mathbb{S}^p \subset \mathbb{R}^{p+1}$ we have

$$dX_t = -\frac{r-1}{2}\, \frac{X_t}{X_t^T X_t}\, dt + \left(Id - \frac{X_t X_t^T}{X_t^T X_t} \right)\, dB_t \tag{6.27}$$

In terms of the cross product (6.24) we have

$$dX_t = -\frac{r-1}{2}\, \frac{X_t}{X_t^T X_t}\, dt + \frac{X_t}{\sqrt{X_t^T X_t}} \wedge dB_t$$

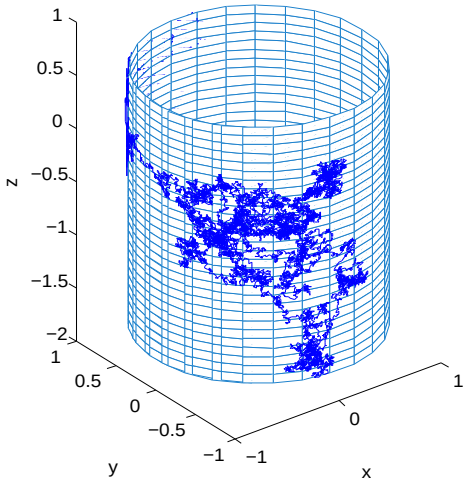
The next picture illustrates a realization of a Brownian motion on the unit sphere.



Using (6.25), the Brownian motion on the cylinder with unit radius is given by

$$\left\{\begin{array}{rcl} dX_t^1 &=& -\frac{1}{2}\,\frac{X_t^1}{(X_t^1)^2+(X_t^2)^2}\,dt + \left(\frac{(X_t^2)^2}{(X_t^1)^2+(X_t^2)^2}\,dB_t^1 - \frac{X_t^1X_t^2}{(X_t^1)^2+(X_t^2)^2}\,dB_t^2\right) \\ dX_t^2 &=& -\frac{1}{2}\,\frac{X_t^2}{(X_t^1)^2+(X_t^2)^2}\,dt + \left(-\frac{X_t^1X_t^2}{(X_t^1)^2+(X_t^2)^2}\,dB_t^1 + \frac{(X_t^1)^2}{(X_t^1)^2+(X_t^2)^2}\,dB_t^2\right) \\ dX_t^3 &=& dB_t^3 \end{array}\right. \tag{6.28}$$

The next picture illustrates a realization of a Brownian motion on the cylinder of unit radius.



6.2.4 The infinitesimal generator

Recalling that $\pi = \pi^T$, for any $1 \le k \le r$ we have

$$\begin{aligned} dX_t^k &= -\frac{1}{2}\,\mathbb{H}^k(X_t)\,dt + \sum_{1\le j\le r} \pi_j^k(X_t)\,dB_t^j \\ &= \sum_{1\le j\le r}\left[\frac{1}{2}\,\partial_{\pi_j}(\pi_j^k)(X_t)\,dt + \pi_j^k(X_t)\,dB_t^j\right] \end{aligned}$$

Notice that

$$dX_t^k dX_t^l \sum_{1\leq i,j\leq r} \pi_i^k(X_t)\pi_j^l(X_t)\ dB_t^i dB_t^j \simeq \underbrace{\sum_{1\leq i\leq r} (\pi_i^k\pi_i^l)(X_t)}_{:=(\pi\pi^T)_l^k}\ dt = \pi_l^k(X_t)\ dt$$

Using Ito formula, for any smooth function F on $\mathbb{R}^r$ we have

$$\begin{aligned} dF(X_t) &= \sum_{1\leq k\leq r} \partial_{x_k}(F)(X_t)\ dX_t^k + \frac{1}{2}\sum_{1\leq k,l\leq r} \partial_{x_k,x_l}(F)(X_t)\ dX_t^k dX_t^l \\ &= \langle \partial F(X_t), dX_t\rangle + \frac{1}{2}\mathrm{tr}\left(\partial^2 F(X_t)\ dX_t dX_t^T\right) \\ &= \langle \partial F(X_t), dX_t\rangle + \frac{1}{2}\mathrm{tr}\left(\pi(X_t)\partial^2 F(X_t)\right)\ dt = L(F)(X_t)\ dt + dM_t(F) \end{aligned}$$

with the infinitesimal generator

$$\begin{aligned} L(F) &= \frac{1}{2}\left[\mathrm{tr}\left(\pi\partial^2 F\right) - \partial_{\mathbb{H}}F\right] = \frac{1}{2}\sum_{1\leq j\leq r}\partial^2_{\pi_j}F \\ &= \frac{1}{2}\ \Delta(F) = \frac{1}{2}\ \mathrm{tr}\left(\nabla^2 F\right) \quad (\Longleftarrow \text{ (6.18) \& (6.19)}) \end{aligned}$$

and the martingale $M_t(F)$ given by

$$\begin{aligned} dM_t(F) &= \langle(\partial F)(X_t), \pi(X_t)dB_t\rangle = \langle \pi(X_t)(\partial F)(X_t), dB_t\rangle \\ &= \langle \pi(X_t)\partial F(X_t), dB_t\rangle = \langle(\nabla F)(X_t), dB_t\rangle \\ &= \sum_{1\leq j\leq r}\left[\sum_{1\leq k\leq r}\pi_j^k\ \partial_{x_k}(F)\right](X_t)\ dB_t^j = \sum_{1\leq j\leq r}\partial_{\pi_j}(F)(X_t)\ dB_t^j \end{aligned}$$

Using (6.20), we check that $X_t \in S$ for any t, as soon as $X_0 \in S$. More precisely, we have

$$\begin{aligned} F=\varphi_i \ &\Rightarrow\ \nabla F = 0 \\ &\Rightarrow\ dM_t(F) = \langle(\nabla F)(X_t), dB_t\rangle = 0\ \ \&\ \ L(F) = \frac{1}{2}\ \mathrm{tr}\left(\nabla(\nabla F)\right) = 0 \\ &\Rightarrow\ \ \varphi_i(X_t) = \varphi(X_0) = 0 \end{aligned}$$

Thus, (6.27) can be rewritten as follows

$$dX_t = -\frac{r-1}{2}\ X_t\ dt + \left(Id - X_t X_t^T\right)\ dB_t$$

Monte Carlo simulation

In practice, the sampling of the diffusion process (6.29) requires some discrete time approximation. For instance, an Euler type approximation on a time mesh $(t_n)_{n\geq 0}$ with $(t_n - t_{n-1}) = \epsilon \simeq 0$ is given by the equation

$$X_{t_n}^\epsilon - X_{t_{n-1}}^\epsilon = -\frac{1}{2}\ \mathbb{H}(X_{t_{n-1}}^\epsilon)\ (t_n - t_{n-1}) + \pi(X_{t_{n-1}}^\epsilon)\ \sqrt{t_n - t_{n-1}}\ \overline{B}_n$$

where B_n stands for a sequence of i.i.d. centered and normalized Gaussian r.v. on $\mathbb{R}^r$. Unfortunately any type of these scheme ensure that $X^\epsilon_{t_n}$ stay in the Manifold S. As for deterministic dynamical systems, we often handle this issue by projecting each step on the manifold

$$X^\epsilon_{t_n} = \mathrm{proj}_S\left(X^\epsilon_{t_{n-1}} - \frac{1}{2}\,\mathbb{H}(X^\epsilon_{t_{n-1}})\;(t_n - t_{n-1}) + \pi(X^\epsilon_{t_{n-1}})\;\sqrt{t_n - t_{n-1}}\;\overline{B}_n\right)$$

Another strategy is to use a description of the stochastic process in some judicious chart space. Manifold parametrizations and chart spaces are discussed in section 6.3. We also refer the reader to section 6.4 for an overview of stochastic calculus on chart spaces, with several illustrations.

6.2.5 Stratonovitch differential calculus

We recall that an r-dimensional stochastic differential equation of the form

$$dX_t = b(X_t)\;dt + \sigma(X_t)\;dB_t$$

can be rewritten as a Stratonovitch differential equation

$$\partial X_t = \left[b - \frac{1}{2}\sum_{1\leq j\leq r}\partial_{\sigma_j}\,(\sigma_j)^T\right](X_t)\;\;\partial t + \sigma(X_t)\;\partial B_t \tag{6.29}$$

with the vector fields σ_j on $\mathbb{R}^r$ defined by the column vectors

$$\forall 1\leq j\leq r\qquad \sigma_j := \begin{bmatrix}\sigma_j^1\\ \vdots\\ \sigma_j^r\end{bmatrix}\;\Longrightarrow\;\partial_{\sigma_j}\,(\sigma_j)^T = \begin{bmatrix}\partial_{\sigma_j}\sigma_j^1\\ \vdots\\ \partial_{\sigma_j}\sigma_j^r\end{bmatrix}$$

In other words, we have

$$\forall 1\leq k\leq r\qquad \partial X_t^k = \left[b^k - \frac{1}{2}\sum_{1\leq j\leq r}\partial_{\sigma_j}\sigma_j^k\right](X_t)\;\;\partial t + \sum_{1\leq j\leq r}\sigma_j^k(X_t)\partial B_t^j$$

An heuristic but constructive derivation of these formulae is given below. The Stratonovich and the Ito increments are connected by

$$b^k(X_t)\;\partial t = b^k\left(X_t + \frac{1}{2}\;dX_t\right)\times dt\quad\text{and}\quad\sigma_j^k(X_t)\;\partial B_t^j = \sigma_j^k\left(X_t + \frac{1}{2}\;dX_t\right)\times dB_t^j$$

with the middle state of the increment of X_t given by

$$\frac{X_t + X_{t+dt}}{2} := X_t + \frac{1}{2}\;dX_t$$

Using this rule, we have

$$\begin{aligned}
b^k(X_t)\;\partial t + \sum_{1\leq j\leq r}\sigma_j^k(X_t)\;\partial B_t^j \;&\simeq\; b^k\left(X_t + \frac{1}{2}\;dX_t\right)\times dt + \sum_{1\leq j\leq r}\sigma_j^k\left(X_t + \frac{1}{2}\;dX_t\right)\times dB_t^j\\
&=\; b^k(X_t)\times dt + \sum_{1\leq j\leq r}\sigma_j^k(X_t)\;dB_t^j\\
&\qquad\qquad + \frac{1}{2}\sum_{1\leq j\leq r}\sum_{1\leq i\leq r}\left(\partial_{x_i}\sigma_j^k\right)(X_t)\times dX_t^i dB_t^j + \ldots\\
&=\; b^k(X_t)\times dt + \sum_{1\leq j\leq r}\sigma_j^k(X_t)\;dB_t^j + \frac{1}{2}\sum_{1\leq j\leq r}\left(\partial_{\sigma_j}\sigma_j^k\right)(X_t)\;\;dt
\end{aligned}$$

The last assertion comes from the fact that

$$dX_t^i dB_t^j = \sum_{1\leq l\leq r} \sigma_l^i(X_t)\ dB_t^l dB_t^j = \sigma_j^i(X_t)\ dt$$

Thus, the Stratonovitch formulation of (6.29) is given by

$$\partial X_t = \pi(X_t)\ \partial B_t$$

In much the same way, using the fact that

$$\begin{aligned}\left(\partial_{\pi_j} F\right)(X_t)\ \partial B_t^j &= \left(\partial_{\pi_j} F\right)\left(X_t + \frac{1}{2}\ dX_t\right)\ dB_t^j \\ &= \left(\partial_{\pi_j} F\right)(X_t)\ \ dB_t^j + \frac{1}{2}\sum_{1\leq l\leq r} \partial_{x_l}\left(\partial_{\pi_j} F\right)(X_t)\ \ dX_t^l dB_t^j\end{aligned}$$

and

$$\begin{aligned}\sum_{1\leq l\leq r} \partial_{x_l}\left(\partial_{\pi_j} F\right)(X_t)\ \ dX_t^l dB_t^j &= \sum_{1\leq l\leq r} \partial_{x_l}\left(\partial_{\pi_j} F\right)(X_t) \sum_{1\leq m\leq r} \pi_m^l(X_t)\ dB_t^m dB_t^j \\ &= \left[\sum_{1\leq l\leq r} \pi_l^j(X_t)\ \partial_{x_l}\left(\partial_{\pi_j} F\right)(X_t)\right]\ dt = \partial_{\pi_j}^2 F(X_t)\ dt\end{aligned}$$

we prove that

$$\left(\partial_{\pi_j} F\right)(X_t)\ \partial B_t^j = \left(\partial_{\pi_j} F\right)(X_t)\ \ dB_t^j + \frac{1}{2}\ \left(\partial_{\pi_j}^2 F\right)(X_t)\ dt$$

and therefore

$$\left(\partial_{\pi_j} F\right)(X_t)\ \ dB_t^j = \left(\partial_{\pi_j} F\right)(X_t)\ \partial B_t^j - \frac{1}{2}\ \left(\partial_{\pi_j}^2 F\right)(X_t)\ dt$$

Therefore, the Stratonovitch formulation of the equation

$$dF(X_t) = L(F)(X_t)\ dt + \sum_{1\leq j\leq r} \left(\partial_{\pi_j} F\right)(X_t)\ dB_t^j$$

is given by

$$\begin{aligned}\partial F(X_t) &= \left[L(F) - \frac{1}{2}\sum_{1\leq j\leq r} \partial_{\pi_j}^2 F\right](X_t)\ dt + \sum_{1\leq j\leq r} \left(\partial_{\pi_j} F\right)(X_t)\ \partial B_t^j \\ &= \sum_{1\leq j\leq r} \left(\partial_{\pi_j} F\right)(X_t)\ \partial B_t^j = \sum_{1\leq k\leq r} \left(\partial_{x_k} F\right)(X_t) \sum_{1\leq j\leq r} \pi_j^k(X_t)\ \partial B_t^j \\ &= \sum_{1\leq k\leq r} \partial_{x_k} F(X_t)\ \partial X_t^k = \langle (\partial F)(X_t), \partial X_t\rangle\end{aligned}$$

6.2.6 Projected diffusions on manifolds

We let $X_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^r \end{pmatrix}$ be the $\mathbb{R}^r$-valued diffusion defined by

$$\begin{aligned} dX_t &= \pi(X_t)\ (b(X_t)dt + \sigma(X_t)\ dB_t) - \frac{1}{2}\ \mathbb{H}_\sigma(X_t)\ dt \\ &= \overline{b}(X_t)dt + \left[\overline{\sigma}(X_t)\ dB_t - \frac{1}{2}\ \mathbb{H}_\sigma(X_t)\ dt\right] \end{aligned} \tag{6.30}$$

with

$$\overline{\sigma}(x) = \pi(x)\sigma(x) \quad \text{and} \quad \overline{b}(x) = \pi(x)b(x)$$

where B_t stands for a standard r-dimensional Brownian motion and

$$\sigma = \begin{bmatrix} \sigma_1^1 & \dots & \sigma_r^1 \\ \vdots & \vdots & \vdots \\ \sigma_1^r & \dots & \sigma_r^r \end{bmatrix} \quad \text{and} \quad \mathbb{H}_\sigma(x) = -\begin{bmatrix} \sum_{1\le j\le r} \partial_{\overline{\sigma}_j}\overline{\sigma}_j^1 \\ \vdots \\ \sum_{1\le j\le r} \partial_{\overline{\sigma}_j}\overline{\sigma}_j^r \end{bmatrix}$$

Using (6.29), the Stratonovitch formulation of the above equation is given by

$$\partial X_t = \overline{b}(X_t)\ \partial t + \overline{\sigma}(X_t)\ \partial B_t$$

In this situation, we have

$$dX_t dX_t^T = \overline{\sigma}(X_t)\ dB_t\ dB_t^T \overline{\sigma}(X_t)^T = \left(\overline{\sigma}\overline{\sigma}^T\right)(X_t)$$

This yields, for any smooth function F on $\mathbb{R}^r$ the Ito formula.

$$\begin{aligned} dF(X_t) &= \langle \partial F(X_t), dX_t\rangle + \frac{1}{2}\mathrm{tr}\left(\partial^2 F(X_t)\ dX_t dX_t^T\right) \\ &= \langle \partial F(X_t), dX_t\rangle + \frac{1}{2}\mathrm{tr}\left(\left(\overline{\sigma}\overline{\sigma}^T\right)(X_t)\partial^2 F(X_t)\right)\ dt = L(F)(X_t)\ dt + dM_t(F) \end{aligned}$$

with

$$L(F) = \partial_{\overline{b}} F + \frac{1}{2}\left[\mathrm{tr}\left(\left(\overline{\sigma}\overline{\sigma}^T\right)\partial^2 F\right) - \ \partial_{\mathbb{H}_\sigma} F\right]$$

and the martingale

$$dM_t(F) = \langle \partial F(X_t), \overline{\sigma}(X_t)dB_t\rangle = \langle \nabla F(X_t), \sigma(X_t)dB_t\rangle$$

To get one step further in our discussion, we notice that

$$\begin{aligned}
&\mathrm{tr}\left(\left(\overline{\sigma}\overline{\sigma}^T\right)(x)\partial^2 F(x)\right)\\
&= \sum_{1\le k,l\le r}\ \partial_{x_k,x_l}(F)(x)\ \left(\overline{\sigma}(x)\overline{\sigma}^T(x)\right)_l^k = \sum_{1\le j,k,l\le r}\ \overline{\sigma}_j^k(x)\overline{\sigma}_j^l(x)\ \partial_{x_k,x_l}(F)(x)\\
&= \sum_{1\le j\le r}\ \sum_{1\le k\le r}\overline{\sigma}_j^k(x)\ \partial_{x_k}\left(\sum_{1\le l\le r}\overline{\sigma}_j^l\ \partial_{x_l}F\right)(x)\\
&\qquad\qquad -\sum_{1\le l\le r}\ \left\{\sum_{1\le j\le r}\ \left[\sum_{1\le k\le r}\overline{\sigma}_j^k(x)\ \partial_{x_k}\left(\overline{\sigma}_j^l\right)(x)\right]\right\}\ \partial_{x_l}F\\
&= \sum_{1\le j\le r}\ \partial_{\overline{\sigma}_j}\left(\partial_{\overline{\sigma}_j}F\right)(x) + \partial_{\mathbb{H}_\sigma}F(x)
\end{aligned}$$

This implies that

$$L(F) \;=\; \partial_{\overline{b}}(F) + \frac{1}{2}\sum_{1\le j\le r}\partial^2_{\overline{\sigma}_j}(F) = \langle b,\nabla F\rangle + \frac{1}{2}\,\mathrm{tr}\left(\sigma^T\nabla\ \sigma^T\nabla F\right) \tag{6.31}$$

The r.h.s. formulation comes from the fact that

$$\overline{\sigma}^T\partial F = \begin{pmatrix} \sum_{1\le k\le r}\sigma_1^k\partial_{x_k}F\\ \vdots\\ \sum_{1\le k\le r}\sigma_r^k\partial_{x_k}F\end{pmatrix}$$

$$\begin{aligned}
\Longrightarrow\ \partial\left(\overline{\sigma}^T\partial F\right) &= \left[\sum_{1\le k\le r}\partial\left(\sigma_1^k\partial_{x_k}F\right),\ldots,\sum_{1\le k\le r}\partial\left(\sigma_r^k\partial_{x_k}F\right)\right]\\
&= \begin{bmatrix} \sum_{1\le k\le r}\partial_{x_1}\left(\sigma_1^k\partial_{x_k}F\right) & \cdots & \sum_{1\le k\le r}\partial_{x_1}\left(\sigma_r^k\partial_{x_k}F\right)\\ \vdots & & \vdots\\ \sum_{1\le k\le r}\partial_{x_r}\left(\sigma_1^k\partial_{x_k}F\right) & \cdots & \sum_{1\le k\le r}\partial_{x_r}\left(\sigma_r^k\partial_{x_k}F\right)\end{bmatrix}
\end{aligned}$$

$$\Longrightarrow \overline{\sigma}^T\left(\partial\left(\overline{\sigma}^T\partial F\right)\right) = \begin{bmatrix} \sum_{1\le k,l\le r}\sigma_1^l\partial_{x_l}\left(\sigma_1^k\partial_{x_k}F\right) & \cdots & \sum_{1\le k,l\le r}\sigma_1^l\partial_{x_l}\left(\sigma_r^k\partial_{x_k}F\right)\\ \vdots & & \vdots\\ \sum_{1\le k,l\le r}\sigma_r^l\partial_{x_l}\left(\sigma_1^k\partial_{x_k}F\right) & \cdots & \sum_{1\le k,l\le r}\sigma_r^l\partial_{x_l}\left(\sigma_r^k\partial_{x_k}F\right)\end{bmatrix}$$

This yields

$$\mathrm{tr}\left(\overline{\sigma}^T\partial\ \overline{\sigma}^T\partial F\right) = \mathrm{tr}\left(\sigma^T\nabla\ \sigma^T\nabla F\right) = \sum_{1\le j\le r}\ \sum_{1\le k,l\le r}\sigma_j^l\partial_{x_l}\left(\sigma_j^k\partial_{x_k}F\right) = \sum_{1\le j\le r}\partial^2_{\overline{\sigma}_i}(F)$$

Using (6.20), we check that $X_t\in S$ for any t, as soon as $X_0\in S$. More precisely, we have

$$\begin{aligned}
F=\varphi_i \;&\Rightarrow\; \nabla F = 0\\
&\Rightarrow\; \left\{\begin{array}{rcl} dM_t(F) &=& \langle(\nabla F)(X_t),\sigma(X_t)dB_t\rangle = 0\\ L(F) &=& \langle b,\nabla F\rangle + \frac{1}{2}\,\mathrm{tr}\left(\sigma^T\nabla\ \sigma^T\nabla F\right) = 0\end{array}\right.\\
&\Rightarrow\; \varphi_i(X_t) = \varphi(X_0) = 0
\end{aligned}$$

6.2.7 Brownian motion on orbifolds

We let $S = \varphi^{-1}(0) \subset \mathbb{R}^{r=p+1}$ be some hypersurface, and $\mathcal{H}$ be a subgroup of the orthogonal group $\mathbb{O}(r)$ on $\mathbb{R}^r$, acting on S, such that

$$\forall h \in \mathcal{H} \quad \forall x \in S \quad hx \in S \quad (\Longrightarrow \varphi(x) = \varphi(hx))$$

The prototype of model we have in head is the unit sphere

$$S = \mathbb{S}^p = \left\{ x = (x_i)_{1\leq i\leq r}^T \ : \ \varphi(x) := \sqrt{\sum_{1\leq i\leq r} x_i^2} - 1 = 0 \right\} \subset \mathbb{R}^{p+1}$$

and the group action induced by the subgroup

$$\mathcal{O} := \left\{ h = \begin{pmatrix} \epsilon_1 & 0 & \dots & 0 \\ 0 & \epsilon_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 & \dots & \epsilon_r \end{pmatrix} \ : \ \forall 1 \leq i \leq r \quad \epsilon_i \in \{-1, 1\} \right\} \tag{6.32}$$

In the case of the sphere the quotient manifold is isomorphic to the positive orthant $S/\mathcal{H} = S \cap \mathbb{R}^r_+$

$$\langle hx, hy\rangle = x^T h^T h y = x^T y = \langle x, y\rangle$$

Quotient manifolds defined by the orbit space $S/\mathcal{H}$ are often called orbifolds.

By construction, we have

$$\begin{aligned} \varphi_h(x) := \varphi(hx) \quad \Longrightarrow \quad \partial_{x_i}(\varphi_h)(x) &= \textstyle\sum_{1\leq k\leq r} (\partial_{x_k}\varphi)(hx)\ \partial_{x_i}\left(\sum_{1\leq j\leq r} h_k^j\ x_j\right) \\ &= \textstyle\sum_{1\leq k\leq r}\ h_k^j\ (\partial_{x_k}\varphi)(hx) = \sum_{1\leq k\leq r}\ (h^T)_j^k\ (\partial_{x_k}\varphi)(hx) \end{aligned}$$

$$\Longrightarrow \quad (\partial\varphi_h)(x) = h^{-1}(\partial\varphi)(hx) = h^T(\partial\varphi)(hx)$$

$$\begin{aligned} \Longrightarrow \quad \|(\partial\varphi_h)(x)\|^2 &= \langle(\partial\varphi_h)(x), (\varphi_h)(x)\rangle \\ &= (\partial\varphi)(hx)^T hh^T(\partial\varphi)(hx) = (\partial\varphi)(hx)^T(\partial\varphi)(hx) \\ &= \|(\partial\varphi)(hx)\|^2 \end{aligned}$$

This shows that the unit normal at hx is given by

$$h\ \frac{(\partial\varphi_h)(x)}{\|(\partial\varphi_h)(x)\|} = \frac{(\partial\varphi)(hx)}{\|(\partial\varphi)(hx)\|}$$

On the other hand, under our assumptions we have

$$\varphi_h(x) := \varphi(hx) = \varphi(x) \Longrightarrow \frac{(\partial\varphi_h)(x)}{\|(\partial\varphi_h)(x)\|} = \frac{(\partial\varphi)(x)}{\|(\partial\varphi)(x)\|}$$

This implies that

$$h\ \frac{(\partial\varphi_h)(x)}{\|(\partial\varphi_h)(x)\|} = h\ \frac{(\partial\varphi)(x)}{\|(\partial\varphi)(x)\|} = \frac{(\partial\varphi)(hx)}{\|(\partial\varphi)(hx)\|}$$

from which we prove that

$$\begin{aligned}
&\pi(x) = Id - \frac{(\partial\varphi)(x)}{\|(\partial\varphi)(hx)\|}\frac{(\partial\varphi)(x)^T}{\|(\partial\varphi)(x)\|} \\
&\implies h\pi(x)h^T = Id - h\,\frac{(\partial\varphi)(x)}{\|(\partial\varphi)(x)\|}\left(h\,\frac{(\partial\varphi)(x)}{\|(\partial\varphi)(x)\|}\right)^T \qquad (6.33)\\
&\qquad = Id - \frac{(\partial\varphi)(hx)}{\|(\partial\varphi)(hx)\|}\frac{(\partial\varphi)^T(hx)}{\|(\partial\varphi)(hx)\|} \implies h\pi(x)h^T = \pi(hx)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\frac{(\partial\varphi_h)}{\|(\partial\varphi_h)\|}(x) = h^T\frac{(\partial\varphi)(hx)}{\|(\partial\varphi)(hx)\|} = \frac{\partial\varphi}{\|(\partial\varphi)\|}(x)\\
&\Rightarrow\ \partial_{x_m}\left[\frac{(\partial_{x_m}\varphi_h)(x)}{\|(\partial\varphi_h)(x)\|}\right] = \textstyle\sum_{1\le j\le r} h_j^m\ \partial_{x_m}\left[\frac{(\partial_{x_j}\varphi)(hx)}{\|(\partial\varphi)(hx)\|}\right] = \partial_{x_m}\left[\frac{(\partial_{x_m}\varphi)}{\|(\partial\varphi)\|}\right](x)\\
&\qquad = \sum_{1\le j\le r} h_j^m \sum_{1\le k\le r}\partial_{x_k}\left[\frac{(\partial_{x_j}\varphi)}{\|(\partial\varphi)\|}\right](hx)\ \partial_{x_m}\left(\sum_{1\le i\le r} h_k^i x_i\right)\\
&\qquad = \sum_{1\le j\le r} h_j^m h_k^m \sum_{1\le k\le r}\partial_{x_k}\left[\frac{(\partial_{x_j}\varphi)}{\|(\partial\varphi)\|}\right](hx)
\end{aligned}$$

Taking the sum over m, this implies that

$$\begin{aligned}
&\sum_{1\le j\le r}\overbrace{\left[\sum_{1\le m\le r} h_j^m (h^T)_m^k\right]}^{=1_{j=k}}\ \sum_{1\le k\le r}\partial_{x_k}\left[\frac{(\partial_{x_j}\varphi)}{\|(\partial\varphi)\|}\right](hx) = \sum_{1\le m\le r}\partial_{x_m}\left[\frac{(\partial_{x_m}\varphi)}{\|(\partial\varphi)\|}\right](x) \qquad (6.34)\\
&\implies \sum_{1\le m\le r}\partial_{x_m}\left[\frac{(\partial_{x_m}\varphi)}{\|(\partial\varphi)\|}\right](hx) = \sum_{1\le m\le r}\partial_{x_m}\left[\frac{(\partial_{x_m}\varphi)}{\|(\partial\varphi)\|}\right](x)
\end{aligned}$$

Combining this result with (6.21) we conclude that

$$\begin{aligned}
h\ \mathbb{H}(x) &= \left[\sum_{1\le m\le r}\partial_{x_m}\left(\frac{\partial_{x_m}\varphi}{\|\partial\varphi\|}\right)(x)\right]\ h\left[\frac{\partial\varphi}{\|\partial\varphi\|}(x)\right]\\
&= \left[\sum_{1\le m\le r}\partial_{x_m}\left(\frac{\partial_{x_m}\varphi}{\|\partial\varphi\|}\right)(hx)\right]\ \frac{(\partial\varphi)(hx)}{\|(\partial\varphi)(hx)\|} \quad\Rightarrow\quad h\ \mathbb{H}(h^Tx) = \mathbb{H}(x)
\end{aligned}$$

We let X_t be the Brownian motion on S defined in (6.29). Using (6.33) and (6.34), for any $h\in\mathcal{H}$ we have

$$\begin{aligned}
Y_t = hX_t \Rightarrow dY_t = hdX_t &= -\frac{1}{2}\ h\mathbb{H}(h^T(hX_t))\ dt + h\pi(X_t)h^T\ hdB_t\\
&= -\frac{1}{2}\ \mathbb{H}(Y_t)\ dt + \pi(Y_t)\ dB_t^{(h)} \qquad (6.35)
\end{aligned}$$

where $B_t^{(h)}$ is a standard r-dimensional Brownian motion. Roughly speaking, this result shows that all the stochastic processes in the same orbit

$$\mathrm{Orb}_{\mathcal{H}}(X) = \{(hX_t)_{t\geq 0} \ : \ h \in \mathcal{H}\}$$

only differ by changing their driving Brownian motion. This coupling technique allows to define in a unique way the Brownian motion on the quotient manifold $S/\mathcal{H}$.

6.3 Parametrizations and Charts

We denote by

$$\psi \ : \ \theta \in S_\psi \subset \mathbb{R}^p \mapsto \psi(\theta) = \left(\psi^1(\theta), \ldots, \psi^r(\theta)\right)^T \in S \subset \mathbb{R}^r \tag{6.36}$$

a given smooth parametrization of S, with a well defined smooth inverse mapping

$$\phi = \psi^{-1} \ : \ x \in S \mapsto \phi(x) = (\phi_1(x), \ldots, \phi_p(x))^T \in S_\phi \subset \mathbb{R}^p$$

To clarify the presentation, we further assume that the manifold S can be parametrized by a single map ψ, and thus thus with a single chart coordinate.

By construction, we have

$$\forall 1 \leq l \leq q \quad \forall \theta \in S_\psi \qquad \varphi_l(\psi(\theta)) = 0$$

$$\Downarrow$$

$$\partial_{\theta^j} \left(\varphi_l \circ \psi\right)(\theta) = \sum_{1\leq k\leq r} \left(\partial_{x_k}\varphi_l\right)(\psi(\theta)) \ \partial_{\theta_i}\psi^l(\theta) = \langle \left(\partial\varphi_l\right)(\psi(\theta)), \partial_{\theta_i}\psi(\theta)\rangle = 0$$

$$\Downarrow$$

$$\forall x \in S \qquad \langle \partial\varphi_l(x), \left(\partial_{\theta_i}\psi\right)(\phi(x))\rangle = 0$$

To clarify the presentation, we set

$$\left(\partial_{\theta_i}\psi\right)_\phi \ : \ x \in S \mapsto \left(\partial_{\theta_i}\psi\right)_\phi(x) := \left(\partial_{\theta_i}\psi\right)(\phi(x)) \in T_x(S)$$

and

$$\left(\partial\phi^i\right)_\psi \ : \ \theta \in S_\psi \mapsto \left(\partial\phi^i\right)_\psi(\theta) = \left(\partial\phi^i\right)(\psi(\theta)) \in \mathbb{R}^r$$

In this notation, we have shown that

$$T(S) = \mathrm{Vect}\left(\left(\partial_{\theta_1}\psi\right)_\phi, \ldots, \left(\partial_{\theta_p}\psi\right)_\phi\right)$$

in the sense that

$$\forall x \in S \qquad T_x(S) = \mathrm{Vect}\left(\left(\partial_{\theta_1}\psi\right)_\phi(x), \ldots, \left(\partial_{\theta_p}\psi\right)_\phi(x)\right)$$

We let e_i be the unit vectors on $S_\psi(\subset \mathbb{R}^p)$ defined by

$$\forall 1 \leq i \leq p \qquad e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{-th term} \tag{6.37}$$

We end this section with an alternative description of the vector fields $(\partial_{\theta_i}\psi)_\phi$. For each $1 \leq i \leq p$, we let $c_i(t) = \phi(x) + t\ e_i$ be a curve in S_ψ with starting point $\phi(x)$ and velocity e_i and

$$C_i(t) = \psi\left(\phi(x) + t\ e_i\right)$$

is the pushed forward curve in the manifold S with velocity

$$\frac{dC_i}{dt}(0) = \sum_{1\leq j\leq p} \left(\partial_{\theta_j}\psi\right)_\phi(x)\ e_i^j = \left(\partial_{\theta_i}\psi\right)_\phi(x)$$

In differential geometry these vector fields in $T_x(S)$ are often denoted using the somehow misleading notation

$$\frac{\partial}{\partial\theta_i}\Big|_x := \left(\partial_{\theta_i}\psi\right)_\phi(x) \tag{6.38}$$

The main reason for this pretty strange notation will become clear in section 6.3.2 (see formula (6.46)).

6.3.1 Orthogonal projection operators

We let $g = (g_{i,j})_{1\leq i,j\leq p}$ be the $(p \times p)$-matrix field on S_ψ defined by

$$\forall 1 \leq i,j \leq p \qquad g_{i,j} := \left\langle \partial_{\theta_i}\psi, \partial_{\theta_j}\psi \right\rangle \qquad \left(\Leftrightarrow \forall\theta \in S_\psi \quad g_{i,j}(\theta) := \left\langle \partial_{\theta_p}\psi(\theta), \partial_{\theta_p}\psi(\theta) \right\rangle\right) \tag{6.39}$$

We also consider the push-forward matrix fields on S given by

$$g_\phi = (g_{\phi,i,j})_{1\leq i,j\leq p} = (g_{i,j} \circ \phi)_{1\leq i,j\leq p}$$

By construction, the projection of any vector field W on $\mathbb{R}^r$ onto $T(S)$ is given by

$$\pi(W) = \sum_{1\leq i\leq p} \left\langle \sum_{1\leq i\leq p} g_\phi^{i,j}\ \left(\partial_{\theta_j}\psi\right)_\phi, W \right\rangle\ \left(\partial_{\theta_i}\psi\right)_\phi \quad \text{with} \quad g^{-1} = (g^{i,j})_{1\leq i,j\leq p}$$

in the sense that

$$\pi(x)(W(x)) = \sum_{1\leq i\leq p} \left\langle \sum_{1\leq i\leq p} g_\phi^{i,j}(x)\ \left(\partial_{\theta_j}\psi\right)_\phi(x), W(x) \right\rangle\ \left(\partial_{\theta_i}\psi\right)_\phi(x)$$

We let $W_\psi = W \circ \psi$ be the pull back vector field on the parameter space, and we denote by $\pi_\psi(\theta) = \pi(\psi(\theta)) = \pi(x)$ the orthonormal projection functional onto $T_x(S)$ with $x = \psi(\theta)$. In this notation, we have

$$\pi_\psi(\theta)(W_\psi(\theta)) = \sum_{1\leq i\leq p} \left\langle \sum_{1\leq i\leq p} g^{i,j}(\theta)\ \left(\partial_{\theta_j}\psi\right)(\theta), W_\psi(\theta) \right\rangle\ \left(\partial_{\theta_i}\psi\right)(\theta)$$

or in a more synthetic form

$$\pi_\psi(W_\psi) = \sum_{1\leq i\leq p} \left\langle \sum_{1\leq i\leq p} g^{i,j}\ \partial_{\theta_j}\psi, W_\psi \right\rangle\ \partial_{\theta_i}\psi$$

By construction, for any $1 \leq i, j \leq p$ we also have that

$$\begin{array}{rcl} \phi^i(\psi(\theta)) = \theta^i \quad \Rightarrow \quad \partial_{\theta^j}\left(\phi^i \circ \psi\right))(\theta) & = & \sum_{1\leq k\leq r} \left(\partial_{x_k}\phi^i\right)_\psi(\theta)\ \partial_{\theta^j}\psi^k(\theta) \\ & = & \left\langle \left(\partial\phi^i\right)_\psi(\theta), \partial_{\theta^j}\psi(\theta) \right\rangle = 1_{i=j} \end{array}$$

so that

$$\forall x \in S \qquad \left\langle \left(\partial\phi^i\right)(x), \left(\partial_{\theta^j}\psi\right)_\phi(x) \right\rangle = 1_{i=j}$$

This implies that

$$\begin{array}{rcl} \nabla\phi^i = \pi\left(\partial\phi^i\right) & = & \sum_{1\leq k\leq p} \left\langle \sum_{1\leq l\leq p} g_\phi^{k,l}\ \left(\partial_{\theta^l}\psi\right)_\phi, \partial\phi^i \right\rangle\ \left(\partial_{\theta^k}\psi\right)_\phi \\ & = & \sum_{1\leq k\leq p} g_\phi^{i,k}\ \left(\partial_{\theta^k}\psi\right)_\phi \end{array} \tag{6.40}$$

and

$$\sum_{1\leq i\leq p} g_{\phi,j,i}\ \nabla\phi^i = \sum_{1\leq k\leq p}\sum_{1\leq i\leq p} g_{\phi,j,i}\ g_\phi^{i,k}\ \left(\partial_{\theta^k}\psi\right)_\phi = \left(\partial_{\theta^j}\psi\right)_\phi$$

By construction, we have

$$\left\langle \left(\nabla\phi^i\right), \left(\partial_{\theta^j}\psi\right)_\phi \right\rangle = 1_{i=j} \quad \text{and} \quad \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle = g_\phi^{i,j} \tag{6.41}$$

We check these claims using the fact that

$$\left\langle \left(\nabla\phi^i\right), \left(\partial_{\theta^j}\psi\right)_\phi \right\rangle = \sum_{1\leq k\leq p} g_\phi^{i,k}\ g_{\phi,k,j} = 1_{i=j}$$

and

$$\begin{array}{rcl} \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle & = & \sum_{1\leq k,l\leq p} g_\phi^{i,k}\ g_\phi^{j,l}\ \left\langle \left(\partial_{\theta^k}\psi\right)_\phi, \left(\partial_{\theta^l}\psi\right)_\phi \right\rangle \\ & = & \sum_{1\leq k\leq p} g_\phi^{i,k} \sum_{1\leq l\leq p} g_\phi^{j,l}\ g_{\phi,l,k} = g_\phi^{i,j} \end{array}$$

In summary, the vector fields $\nabla\phi^i$ form a new basis of $T(S)$

$$T(S) = \text{Vect}\left(\nabla\phi^1, \ldots, \nabla\phi^p\right) \quad \text{with the scalar product} \quad \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle = g_\phi^{i,j}$$

and the change of basis formulae are given by

$$\left(\partial_{\theta^i}\psi\right)_\phi = \sum_{1\leq j\leq p} g_{\phi,i,j}\ \nabla\phi^j \quad \text{and} \quad \nabla\phi^i = \sum_{1\leq j\leq p} g_\phi^{i,j}\ \left(\partial_{\theta^j}\psi\right)_\phi \tag{6.42}$$

Expressed in this new basis vector fields, the orthogonal projection operators π take the form

$$\pi(W) = \sum_{1\leq i\leq p} \left\langle \sum_{1\leq j\leq p} g_{\phi,i,j}\ \nabla\phi^j, W \right\rangle\ \nabla\phi^i$$

Notice that

$$W = \sum_{1\leq i\leq p} V_\phi^i\ \left(\partial_{\theta^i}\psi\right)_\phi \implies \forall 1\leq i\leq p \quad V_\phi^i = \left\langle W, \nabla\phi^i \right\rangle$$

Rewritten in a slightly different form we have

$$\pi(W) = \sum_{1\leq j\leq p} \left\langle \nabla\phi^j, W \right\rangle \sum_{1\leq i\leq p} g_{\phi,j,i}\ \nabla\phi^i = \sum_{1\leq j\leq p} \left\langle \nabla\phi^j, W \right\rangle\ \left(\partial_{\theta^j}\psi\right)_\phi \tag{6.43}$$

6.3.2 Riemannian structures

We consider a smooth curve in the parameter space $c\ :\ t\in[0,1]\mapsto c(t) = \left(c^1(t), \ldots, c^p(t)\right)^T \in S_\psi$ starting at some parameter state $c(0)=\theta\in S_\psi$, with a velocity vector field $V(\in\mathbb{R}^p)$, that is we have that

$$\frac{dc}{dt} = \left(\frac{dc^1}{dt}, \ldots, \frac{dc^p}{dt}\right)^T = V(c(t)) = \left(V^1(c(t)), \ldots, V^p(c(t))\right)^T$$

The function c is called an integral curve of V (a.k.a. V-integral curve). By construction, $C(t) := \psi(c(t))$ is a smooth curve on S and we have

$$\forall 1\leq i\leq q \qquad \frac{dC}{dt} = \frac{d}{dt}\psi(c(t)) = \sum_{1\leq i\leq p} \left(\partial_{\theta_i}\psi\right)(c(t))\ V^i(c(t))$$

For $t=0$, this implies that

$$\frac{dC}{dt}(0) = \sum_{1\leq i\leq p} V_\phi^i(x)\ \left(\partial_{\theta_i}\psi\right)_{\phi(x)}$$

with

$$x = \psi(\theta) \Leftrightarrow \theta = \phi(x) \quad \text{and} \quad V_\phi^i(x) = V^i(\phi(x))$$

In other words, C is an integral curve of the vector field

$$W(x) = \sum_{1\leq i\leq p} V_\phi^i(x)\ \left(\partial_{\theta_i}\psi\right)_{\phi(x)}$$

For any smooth function $F = f \circ \phi$ on S we have

$$\begin{aligned}
\frac{d}{dt}F(C(t)) &= \sum_{1\le k\le r} W^k(C(t))\ (\partial_{x_k}F)(C(t)) = \partial_W(F)(C(t)) \\
&= \sum_{1\le j\le p} V^j(c(t)) \sum_{1\le k\le r} (\partial_{x_k}F)(\psi(c(t)))\ \left(\partial_{\theta_j}\psi^k\right)(c(t)) \\
&= \sum_{1\le j\le p} V^j(c(t))\ \left(\partial_{\theta_j}(F\circ\psi)\right)(c(t)) \\
&= \sum_{1\le j\le p} V^j(c(t))\ \left(\partial_{\theta_j}f\right)(c(t)) = \partial_V(f)(c(t)) = \frac{d}{dt}f(c(t)) \qquad (6.44)
\end{aligned}$$

This shows that

$$\frac{d}{dt}F(C(t)) = \partial_W(F)(C(t)) = \partial_V(f)(c(t)) = \frac{d}{dt}f(c(t)) \quad \stackrel{t=0}{\Longrightarrow} \quad \partial_W F = (\partial_V f)\circ\phi \qquad (6.45)$$

Vector fields can also be interpreted as differential operators

$$W\ :\ F\ \mapsto\ W(F) = \partial_W(F) = W^T\partial F = \langle W, \partial F\rangle$$

In this interpretation, rewritten in terms of (6.38), we have

$$\left(\partial_{\theta_j}(F\circ\psi)\right)_\phi = \sum_{1\le k\le r} \left(\frac{\partial}{\partial\theta_i}\right)^k\ (\partial_{x_k}F) = \left\langle \frac{\partial}{\partial\theta_i}, \partial F\right\rangle = \frac{\partial}{\partial\theta_i}(F)$$

and therefore

$$\forall x\in S \qquad W(x) = \sum_{1\le i\le p} V^i_\phi(x)\ \left(\frac{\partial}{\partial\theta_i}\right)_{|x} \ \Leftrightarrow\ W = \sum_{1\le i\le p} V^i_\phi\ \frac{\partial}{\partial\theta_i} \qquad (6.46)$$

In this synthetic notation, we have

$$W(F) = \sum_{1\le i\le p} V^i_\phi\ \frac{\partial}{\partial\theta_i}(F) = \sum_{1\le i\le p} V^i_\phi\ \partial_{\theta_i}(f) \quad \text{with} \quad f = F\circ\psi \quad \Leftrightarrow F = f\circ\phi$$

This induces a one to one *linear* mapping between the tangent spaces $T_\theta(S_\psi)$ of the parameter space and the tangent space $T_{\psi(\theta)}(S)$ on the manifold S. This mapping is called the pushforward of the vector fields on $T_\theta(S_\psi)$ into $T_{\psi(\theta)}(S)$, and it is given by

$$(d\psi)\ :\ V\in T(S_\psi)\mapsto (d\psi)(V) := \sum_{1\le i\le p} V^i\ (\partial_{\theta_i}\psi)\ \in T_\psi(S)$$

in the sense that

$$(d\psi)_\theta\ :\ V(\theta)\in T(S_\psi)\mapsto (d\psi)_\theta(V(\theta)) := \sum_{1\le i\le p} V^i(\theta)\ (\partial_{\theta_i}\psi)(\theta)\ \in T_{\psi(\theta)}(S)$$

Alternatively, we have

$$(d\psi)_\phi \ : \ V_\phi \in T_\phi(S_\psi) \mapsto (d\psi)_\phi (V_\phi) := \sum_{1\leq i\leq p} V_\phi^i \ (\partial_{\theta_i}\psi)_\phi \ \in T(S)$$

in the sense that

$$(d\psi)_{\phi(x)} \ : \ V(\phi(x)) = V_\phi(x) \in T_{\phi(x)}(S_\psi) \mapsto (d\psi)_{\phi(x)} (V_\phi(x)) := \sum_{1\leq i\leq p} V_\phi^i(x) \ (\partial_{\theta_i}\psi)_\phi (x) \ \in T_x(S)$$

Finally, we notice that

$$W = (d\psi)_\phi(V_\phi) = \sum_{1\leq i\leq p} V_\phi^i \ (\partial_{\theta^i}\psi)_\phi \Rightarrow \left\langle \nabla\phi^j, W \right\rangle = \sum_{1\leq i\leq p} V_\phi^i \left\langle \nabla\phi^j, (\partial_{\theta^i}\psi)_\phi \right\rangle = V_\phi^j$$

Thus, for any $W \in T(S)$ we have

$$V_\phi = (d\psi)_\phi^{-1}(W) = \begin{bmatrix} (\nabla\phi^1)^T \\ \vdots \\ (\nabla\phi^p)^T \end{bmatrix} W$$

The parameter space $S_\psi \subset \mathbb{R}^p$ is free of any constraints and we have

$$T_\theta(S_\psi) = \text{Vect}\,(e_1, \ldots, e_p)$$

with the unit vectors defined in (6.37).

In this notation, $(d\psi)$ maps the basis functions e_i of $T_\theta(S_\psi)$ into the basis functions $(\partial_{\theta_i}\psi)$ of $T_{\psi(\theta)}(S)$; that is, we have that

$$(d\psi)\,(e_i) = (\partial_{\theta_i}\psi) \quad \text{and} \quad (d\psi)_\phi^{-1}\,(\partial_{\theta_i}\psi) = e_i$$

It is also essential to notice that

$$\langle (d\psi)\,(V_1), (d\psi)\,(V_2) \rangle = \sum_{1\leq i\leq p} V_1^i \ g_{i,j} \ V_2^j = V_1^T g \ V_2$$

Thus, if we equip the tangent space $T_\theta(S_\psi)$ with the scalar product

$$\langle V_1, V_2 \rangle_g = \sum_{1\leq i\leq p} g_{i,j} \ V_1^i V_2^j$$

the description of $T(S)$ in the chart ϕ given by

$$(W_k)_\psi = (d\psi)(V_k) \Rightarrow \langle (d\psi)\,(V_1), (d\psi)\,(V_2) \rangle = \sum_{1\leq i\leq p} V_1^i V_2^j \ g_{i,j}$$

In summary, we have

$$\langle V_1, V_2 \rangle_g = \langle (W_1)_\psi, (W_2)_\psi \rangle$$

More formally, the (linear) push forward mappings $(d\psi)_\theta$ are smooth isomorphisms between the inner product spaces $\left(T_\theta(S_\psi), \langle ., . \rangle_{g(\theta)}\right)$ and $\left(T_{\psi(\theta)}(S), \langle ., . \rangle\right)$
The scalar product induced by g on the tangent space $T(S_\psi)$ of the parameter space S_ψ is called the Riemannian scalar product.

6.3.3 First order covariant derivatives

Pushed forward functions

Smooth functions F on S are the push forward of functions f on S_ψ, and inversely functions f on the parameter space are the pull back of functions F on S using the relations

$$F = f \circ \phi \quad \text{and} \quad f = F \circ \psi$$

As a rule, we use the letters F and W to denote functions, and vector fields on S and f, V to denote functions and vector fields on the parameter space S_ψ.
We also denote by $F_\psi = F \circ \psi$, resp. $W_\psi = W \circ \psi$, and $f_\phi = f \circ \phi$, resp. $V_\phi = V \circ \phi$, the pull back of W, resp. F, and V, resp. f, w.r.t. ψ and ϕ.

In this notation, differential of push forward functions are given by

$$\partial_{\theta_i}(F_\psi) = \sum_{1\leq j\leq r} \left(\partial_{x_j} F\right)_\psi \ \partial_{\theta_i}\psi^j = (\partial_{\theta_i}\psi)^T \ (\partial F)_\psi$$

In terms of (6.38)

In much the same way, differential of pull back functions are given by the formula

$$\partial_{x_i}(f_\phi) = \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right)_\phi \ \partial_{x_i}\phi^j \quad \Leftrightarrow \quad (\partial f_\phi)_\psi = \sum_{1\leq j\leq p} \partial_{\theta_j} f \ \left(\partial\phi^j\right)_\psi$$

and therefore

$$\begin{aligned}
(\nabla F)_\psi &= \pi_\psi(\partial F)_\psi = \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right) \ \pi_\psi\left(\partial\phi^j\right)_\psi \\
&= \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right) \ (\nabla\phi^j)_\psi = \sum_{1\leq i\leq p} \left[\sum_{1\leq j\leq p} g^{i,j} \ \left(\partial_{\theta_j} f\right)\right] \ \partial_{\theta^i}\psi \qquad (6.47) \\
&= \sum_{1\leq i\leq p} (\nabla_g(f))^i \ \partial_{\theta^i}\psi = d\psi\left(\nabla_g(f)\right) \qquad (6.48)
\end{aligned}$$

with the vector field $\nabla_g f$ on S_ψ given by

$$\nabla_g f := \begin{bmatrix} \sum_{1\leq j\leq p} g^{1,j} \ \left(\partial_{\theta_j} f\right) \\ \vdots \\ \sum_{1\leq j\leq p} g^{p,j} \ \left(\partial_{\theta_j} f\right) \end{bmatrix} = g^{-1}\partial f \qquad (6.49)$$

The last assertion comes from the change of basis formula (6.42). It is also instructive to observe that

$$\nabla_g f = \sum_{1\leq i\leq p} \langle \sum_{1\leq j\leq p} g^{i,j} \ e_j, \partial f\rangle \ e_i = \sum_{1\leq i\leq p} \left[\sum_{1\leq i\leq p} g^{i,j} \ \partial_{\theta^j} f\right] \ e_i \quad \text{and} \quad \left(\nabla\phi^i\right)^j_\psi = \left(\nabla_g\psi^j\right)^i$$

Furthermore, using (6.41) for any

$$F = f \circ \phi \qquad \text{and} \qquad W = (d\psi)_\phi\left(V_\phi\right)$$

we have

$$\begin{aligned}\langle \nabla F, W\rangle &= \sum_{1\leq i,j\leq p} V_\phi^i \ \left(\partial_{v_j} f\right)_\phi \left\langle \nabla\phi^j, \left(\partial_{\theta_i}\psi\right)_\phi \right\rangle \\ &= \sum_{1\leq i\leq p} V_\phi^i \ \left(\partial_{v_i} f\right)_\phi = \langle (\partial f)_\phi, V_\phi\rangle\end{aligned}$$

In much the same way, we have

$$\begin{aligned}\langle (\nabla F)_\psi, W_\psi\rangle &= \sum_{1\leq i,j\leq p} \left(\nabla_g(f)\right)^i \ V^j \ \langle \partial_{\theta^i}\psi, \partial_{\theta^j}\psi\rangle \\ &= \sum_{1\leq i,j\leq p} g_{i,j} \ \left(\nabla_g(f)\right)^i \ V^j = \langle \nabla_g f, V\rangle_g = \langle \partial f, V\rangle\end{aligned}$$

In terms of directional derivatives, we have

$$\left(\partial_W(F)\right)\circ\psi = \langle (\nabla F)_\psi, W_\psi\rangle = \langle \nabla_g f, V\rangle_g$$

In particular, for any couple of functions $F_1 = f_1\circ\phi$ and $F_2 = f_2\circ\phi$ we have

$$\langle \nabla F_1, \nabla F_2\rangle = \left\langle \left(\nabla_g f_1\right)_\phi, \left(\nabla_g f_2\right)_\phi \right\rangle_{g_\phi} \quad \left(= \left\langle \left(\partial f_1\right)_\phi, \left(\partial f_2\right)_\phi \right\rangle_{g_\phi^{-1}} \right) \tag{6.50}$$

We consider the r coordinate projection mappings

$$\chi^i = \psi^i\circ\phi \ : \ x\in S \ \mapsto \chi^i\left(x\right) = \left(\psi^i\circ\phi\right)\left(x\right) = x_i \in \mathbb{R} \tag{6.51}$$

Applying the above formula to $f=\psi^i$, using (6.48) we find that

$$\nabla \ \chi^i = (d\psi)_\phi \left(\left(\nabla_g\psi^i\right)_\phi\right) = \sum_{1\leq j\leq p} \left(\nabla\phi^j\right)_\psi \ \left(\partial_{v_j}\psi^i\right)$$

On the other hand, by (6.43) we have

$$\begin{aligned}\pi(W) &= \sum_{1\leq j\leq p} \left\langle \nabla\phi^j, W\right\rangle \ \left(\partial_{\theta^j}\psi\right)_\phi \\ &= \begin{bmatrix} \sum_{1\leq j\leq p} \ \left\langle \nabla\phi^j, W\right\rangle \ \left(\partial_{\theta^j}\psi^1\right)_\phi \\ \vdots \\ \sum_{1\leq j\leq p} \ \left\langle \nabla\phi^j, W\right\rangle \ \left(\partial_{\theta^j}\psi^r\right)_\phi \end{bmatrix} = \begin{bmatrix} \left(\nabla \ \chi^1\right)^T \\ \vdots \\ \left(\nabla \ \chi^r\right)^T \end{bmatrix} \ W\end{aligned} \tag{6.52}$$

We also have

$$\begin{bmatrix} \Delta \ \chi^1 \\ \vdots \\ \Delta \ \chi^r \end{bmatrix} = -\mathbb{H}$$

We check this claim using the fact that

$$\begin{aligned}\chi^k\left(x\right) = x_k \Rightarrow \partial^2 \ \chi^k = 0 \Rightarrow \ \Delta \ \chi^k &= \sum_{1\leq l\leq r} \ \partial_{\pi_l}\pi_k^l \quad (\Leftarrow (6.18)) \\ &= -\partial_{\mathbb{H}} \ \chi^k = -\mathbb{H}^k \quad (\Leftarrow (6.19))\end{aligned} \tag{6.53}$$

Pushed forward vector fields

We consider the push forward W_x on $T_x(S)$ of a vector field V on $T(S_\psi)$ given by the formula

$$W(x) = (d\psi)_{\phi(x)}\left(V_\phi(x)\right) := \sum_{1\leq j\leq p} V_\phi^j(x)\ \left(\partial_{\theta_j}\psi\right)_\phi(x)$$

We have

$$V_\phi^j = V^j\circ\phi \Longrightarrow \partial_{x_k}\left(V_\phi^j\right)(x) = \sum_{1\leq l\leq p}\left(\partial_{\theta_l}V^j\right)(\phi(x))\ \left(\partial_{x_k}\phi^l\right)(x)$$

and

$$\partial_{x_k}\left(\left(\partial_{\theta_j}\psi^i\right)_\phi\right)(x) = \sum_{1\leq l\leq p}\left(\partial_{\theta_l,\theta_j}\psi^i\right)(\phi(x))\ \left(\partial_{x_k}\phi^l\right)(x)$$

Rewritten in a more synthetic way we have

$$\begin{aligned}\partial_{x_k}V_\phi^j &= \sum_{1\leq l\leq p}\left(\partial_{\theta_l}V^j\right)_\phi\ \partial_{x_k}\phi^l\\ \partial_{x_k}\left(\partial_{\theta_j}\psi\right)_\phi &= \sum_{1\leq l\leq p}\left(\partial_{\theta_l,\theta_j}\psi\right)_\phi\ \partial_{x_k}\phi^l\end{aligned}$$

This implies that

$$\begin{aligned}\partial_{x_k}W^i &= \sum_{1\leq j\leq p}\left[\partial_{x_k}(V_\phi^j)\ \left(\partial_{\theta_j}\psi^i\right)_\phi + V_\phi^j\ \partial_{x_k}\left(\partial_{\theta_j}\psi^i\right)_\phi\right]\\ &= \sum_{1\leq j,l\leq p}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \left(\partial_{\theta_j}\psi^i\right)_\phi + V_\phi^j\ \left(\partial_{\theta_l,\theta_j}\psi^i\right)_\phi\right]\ \partial_{x_k}\phi^l\end{aligned}$$

In vector form, we have

$$\partial W^i = \sum_{1\leq j,l\leq p}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \left(\partial_{\theta_j}\psi^i\right)_\phi + V_\phi^j\ \left(\partial_{\theta_l,\theta_j}\psi^i\right)_\phi\right]\ \partial\phi^l$$

and

$$\partial W = \sum_{1\leq j,l\leq p}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \partial\phi^l\ \left(\partial_{\theta_j}\psi\right)_\phi^T + V_\phi^j\ \partial\phi^l\ \left(\partial_{\theta_l,\theta_j}\psi\right)_\phi^T\right]$$

This implies that

$$\begin{aligned}\nabla W &= \pi(\partial W)\\ &= \sum_{1\leq j,l\leq p}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \nabla\phi^l\ \left(\partial_{\theta_j}\psi\right)_\phi^T + V_\phi^j\ \nabla\phi^l\ \left(\partial_{\theta_l,\theta_j}\psi\right)_\phi^T\right]\\ &= \sum_{1\leq j,k,l\leq p} g_\phi^{l,k}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \ \left(\partial_{\theta_k}\psi\right)_\phi\ \left(\partial_{\theta_j}\psi\right)_\phi^T + V_\phi^j\ \left(\partial_{\theta_k}\psi\right)_\phi\ \left(\partial_{\theta_l,\theta_j}\psi\right)_\phi^T\right]\end{aligned}$$

Taking the trace, we obtain

$$\begin{aligned}\mathrm{tr}\left(\nabla W\right) &= \sum_{1\leq j,k,l\leq p} g_\phi^{l,k}\left[\left(\partial_{\theta_l}V^j\right)_\phi\ \ g_{\phi,k,j} + V_\phi^j\ \left\langle\left(\partial_{\theta_k}\psi\right)_\phi, \left(\partial_{\theta_l,\theta_j}\psi\right)_\phi\right\rangle\right]\\ &= \sum_{1\leq j\leq p}\left(\partial_{\theta_j}V^j\right)_\phi + \frac{1}{2}\sum_{1\leq j,k,l\leq p} g_\phi^{l,k}\ V_\phi^j\ \left(\partial_{\theta_j}\left\langle\partial_{\theta_k}\psi, \partial_{\theta_l}\psi\right\rangle\right)_\phi\end{aligned}$$

This yields

$$\mathrm{tr}\left(\nabla W\right)_\psi \;=\; \sum_{1\leq j\leq p}\partial_{\theta_j}V^j+\sum_{1\leq j\leq p}V^j\;\frac{1}{2}\sum_{1\leq k,l\leq p}g^{l,k}\;\partial_{\theta_j}g_{k,l}$$

Recalling that

$$\sum_{1\leq k,l\leq p}g^{l,k}\;\partial_{\theta_j}g_{k,l}=\mathrm{tr}\left(g^{-1}\partial_{\theta_j}g\right)=\frac{1}{\det(g)}\;\partial_{\theta_j}\left(\det(g)\right)=\frac{2}{\sqrt{\det(g)}}\;\partial_{\theta_j}\left(\sqrt{\det(g)}\right)$$

we conclude that

$$\mathrm{div}\left(W\right)_\psi:=\mathrm{tr}\left(\nabla W\right)_\psi=\sum_{1\leq j\leq p}\frac{1}{\sqrt{\det(g)}}\;\partial_{\theta_j}\left(\sqrt{\det(g)}\;V^j\right):=\mathrm{div}_g(V)\tag{6.54}$$

Choosing $W=\nabla F=\left(d\psi\right)_\phi\left(\nabla_g(f)\right)_\phi$ we also have that

$$\begin{aligned}\left(\Delta F\right)_\psi &= \mathrm{div}\left(\nabla F\right)_\psi:=\mathrm{tr}\left(\nabla^2F\right)_\psi\\ &= \sum_{1\leq j\leq p}\frac{1}{\sqrt{\det(g)}}\;\partial_{\theta_j}\left(\sqrt{\det(g)}\;\sum_{1\leq i\leq p}g^{j,i}\;\partial_{\theta^i}f\right):=\mathrm{div}_g(\nabla_g f)\end{aligned}\tag{6.55}$$

Directional derivatives

We let W_1,W_2 be a couple of vector fields in $T(S)$. We let V_1,V_2 their pull back vector fields so that

$$W_k\circ\psi=\left(d\psi\right)\left(V_k\right):=\sum_{1\leq j\leq p}V_k^j\;\partial_{\theta_j}\psi$$

for any $k=1,2$. We let C_1 be a W_1-integral curve, that is

$$\frac{dC_1}{dt}(t)=W_1\left(C_1(t)\right)\Rightarrow\frac{d}{dt}F(C_1(t))=\sum_{1\leq k\leq r}W_1^k(C_1(t))\;(\partial_{x_k}F)(C_1(t))=\partial_{W_1}(F)(C_1(t))$$

We recall from (6.44) that

$$\partial_{W_1}(F)\circ\psi=\sum_{1\leq j\leq p}V_1^j\;\partial_{\theta_j}\left(F\circ\psi\right)=\partial_{V_1}(F\circ\psi)=\left\langle\nabla_g(F\circ\psi),V_1\right\rangle_g\tag{6.56}$$

Notice that

$$\begin{aligned}\left(\partial_{W_1}(\partial_{W_2}(F))\right)\circ\psi &= \left(\partial_{W_1}(F_2)\right)\circ\psi\quad\text{with}\quad F_2\circ\psi=\sum_{1\leq j\leq p}V_2^j\;\partial_{\theta_j}\left(F\circ\psi\right)=\partial_{V_2}(F\circ\psi)\\ &= \partial_{V_1}(F_2\circ\psi))=\partial_{V_1}\left(\partial_{V_2}(F\circ\psi)\right)\\ &= \sum_{1\leq i,j\leq p}V_1^i\;\partial_{\theta_i}(V_2^j)\;\partial_{\theta_j}\left(F\circ\psi\right)+\sum_{1\leq i,j\leq p}V_1^iV_2^j\;\partial_{\theta_i,\theta_j}\left(F\circ\psi\right)\end{aligned}\tag{6.57}$$

The directional derivative of the vector field W_2 along the curve C_1 is given by

$$\frac{d}{dt}W_2(C_1(t)) = \begin{bmatrix} \frac{d}{dt}W_2^1(C_1(t)) \\ \vdots \\ \frac{d}{dt}W_2^r(C_1(t)) \end{bmatrix} = \begin{bmatrix} \left(\partial_{W_1}W_2^1\right)(C_1(t)) \\ \vdots \\ \left(\partial_{W_1}W_2^r\right)(C_1(t)) \end{bmatrix} := \partial_{W_1}(W_2)(C_1(t))$$

with

$$\partial_{W_1}(W_2) = \begin{bmatrix} \partial_{W_1}W_2^1 \\ \vdots \\ \partial_{W_1}W_2^r \end{bmatrix} \Rightarrow \partial_{W_1}(W_2)\circ\psi = \begin{bmatrix} \partial_{V_1}(W_2^1\circ\psi) \\ \vdots \\ \partial_{V_1}(W_2^r\circ\psi) \end{bmatrix}$$

Using the fact that

$$\begin{aligned} \partial_{V_1}(W_2^k\circ\psi) &= \sum_{1\leq i\leq p} V_1^i\ \partial_{\theta_i}\left[\sum_{1\leq j\leq p} V_2^j\ \partial_{\theta_j}\psi^k\right] \\ &= \sum_{1\leq i,j\leq p} V_1^i\ \partial_{\theta_i}(V_2^j)\ \partial_{\theta_j}\psi^k + \sum_{1\leq i\leq p} V_1^i\ V_2^j\ \partial_{\theta_i,\theta_j}\psi^k \end{aligned}$$

we conclude that

$$\partial_{W_1}(W_2)\circ\psi = \sum_{1\leq i,j\leq p} V_1^i\ \partial_{\theta_i}(V_2^j)\ \partial_{\theta_j}\psi + \sum_{1\leq i\leq p} V_1^i\ V_2^j\ \partial_{\theta_i,\theta_j}\psi$$

The directional covariant derivative is defined by taking the projection on the tangent space $T(S)$

$$\nabla_{W_1}W_2 = \pi\left(\partial_{W_1}(W_2)\right) = \sum_{1\leq i,j\leq p} V_{1,\phi}^i\ \left(\partial_{\theta_i}V_2^j\right)_\phi\ \left(\partial_{\theta_j}\psi\right)_\phi + \sum_{1\leq i\leq p} V_{\phi,1}^i\ V_{\phi,2}^j\ \pi\left(\left(\partial_{\theta_i,\theta_j}\psi\right)_\phi\right)$$

or equivalently

$$(\nabla_{W_1}W_2)\circ\psi = \pi_\psi\left(\left(\partial_{W_1}(W_2)\right)_\psi\right) = \sum_{1\leq i,j\leq p} V_1^i\ \left(\partial_{\theta_i}V_2^j\right)\ \partial_{\theta_j}\psi + \sum_{1\leq i\leq p} V_1^i\ V_2^j\ \pi_\psi\left(\left(\partial_{\theta_i,\theta_j}\psi\right)\right)$$

with

$$\pi_\psi\left(\left(\partial_{\theta_i,\theta_j}\psi\right)\right) = \sum_{1\leq k\leq p} \underbrace{\left\langle \sum_{1\leq l\leq p} g^{k,l}\ \partial_{\theta_l}\psi, \partial_{\theta_i,\theta_j}\psi \right\rangle}_{:=\Gamma_{i,j}^k}\ \partial_{\theta_k}\psi \tag{6.58}$$

The coordinate functions $\Gamma_{i,j}^k$ are called the Christoffel symbol.

In this notation, we have

$$\begin{aligned}(\nabla_{W_1}W_2)\circ\psi &= \sum_{1\le k\le p}\left[\sum_{1\le i\le p} V_1^i\ \left(\partial_{\theta_i}V_2^k\right)+\sum_{1\le i,j\le p}\Gamma_{i,j}^k\ V_1^i\ V_2^j\right]\ \partial_{\theta_k}\psi \\ &= \sum_{1\le k\le p}\left[\sum_{1\le i\le p} V_1^i\ \left\{\left(\partial_{\theta_i}V_2^k\right)+\sum_{1\le j\le p}\Gamma_{i,j}^k\ V_2^j\right\}\right]\ \partial_{\theta_k}\psi=(d\psi)\left(\nabla_{g,V_1}V_2\right)\end{aligned} \tag{6.59}$$

with the Riemannian directional derivative

$$\nabla_{g,V_1}V_2=\begin{bmatrix}\sum_{1\le i\le p}\ V_1^i\ \left\{\left(\partial_{\theta_i}V_2^1\right)+\sum_{1\le j\le p}\Gamma_{i,j}^1\ V_2^j\right\}\\ \vdots\\ \sum_{1\le i\le p}\ V_1^i\ \left\{\left(\partial_{\theta_i}V_2^p\right)+\sum_{1\le j\le p}\Gamma_{i,j}^p\ V_2^j\right\}\end{bmatrix} \tag{6.60}$$

This shows that

$$\left(\nabla_{W_1}\left(\partial_{\theta_l}\psi\right)_\phi\right)\circ\psi=\sum_{1\le i,k\le p}\Gamma_{i,l}^k\ V_1^i\ \partial_{\theta_k}\psi\implies\left(\nabla_{\left(\partial_{\theta_i}\psi\right)_\phi}\left(\partial_{\theta_j}\psi\right)_\phi\right)\circ\psi=\sum_{1\le k\le p}\Gamma_{i,j}^k\ \partial_{\theta_k}\psi$$

On the other hand, we have

$$\begin{aligned}c_1^i(t)=\phi^i(C_1(t))\Rightarrow\ \dot{c}_1^i\ (t) &= \partial_{W_1}(\phi^i)(C_1(t))=\textstyle\sum_{1\le j\le p}\ V_1^j(c_1(t))\ \underbrace{\partial_{\theta_j}\left(\phi^i\circ\psi\right)(c_1(t))}_{=1_{i=j}}\\ &= V_1^i(c_1(t))\end{aligned}$$

Thus, using (6.59), we find that

$$\begin{aligned}&\left(\nabla_{W_1}W_2\right)(C_1(t))\\ &=\sum_{1\le k\le p}\left[\sum_{1\le i\le p}\ \dot{c}_1^i\ (t)\ \left(\partial_{\theta_i}V_2^k\right)(c_1(t))+\sum_{1\le i,j\le p}\Gamma_{i,j}^k(c_1(t))\ \ \dot{c}_1^i\ (t)\ V_2^j(c_1(t))\right]\ \ (\partial_{\theta_k}\psi)\left(c_1(t)\right)\\ &=\sum_{1\le k\le p}\left[\frac{d}{dt}(V_2^k(c_1(t))+\sum_{1\le i,j\le p}\Gamma_{i,j}^k(c_1(t))\ \ \dot{c}_1^i\ (t)\ V_2^j(c_1(t))\right]\ \ (\partial_{\theta_k}\psi)\left(c_1(t)\right)\end{aligned}$$

In differential geometry, the above formula is sometimes written in terms of the linear differential

operator

$$\begin{aligned}
\frac{DW_2}{dt}(t) &:= \left(\nabla_{\dot{C}_1(t)} W_2\right)(C_1(t)) \\
&= \sum_{1\leq k\leq p}\left[\frac{d}{dt}(V_2^k(c_1(t)) + \sum_{1\leq i,j\leq p}\Gamma_{i,j}^k(c_1(t))\ \dot{c}_1^i\ (t)\ V_2^j(c_1(t))\right]\ (\partial_{\theta_k}\psi)(c_1(t)) \\
&= \sum_{1\leq j\leq p}\left[\frac{d}{dt}(V_2^j(c_1(t))\left(\partial_{\theta_j}\psi\right)(c_1(t)) + V_2^j(c_1(t))\sum_{1\leq i,k\leq p}\Gamma_{i,j}^k(c_1(t))\ \dot{c}_1^i\ (t)\ (\partial_{\theta_k}\psi)(c_1(t))\right] \\
&= \sum_{1\leq j\leq p}\left[\frac{d}{dt}(V_2^j(c_1(t))\left(\partial_{\theta_j}\psi\right)(c_1(t)) + V_2^j(c_1(t))\ \left(\nabla_{\dot{C}_1(t)}(\partial_{\theta_l}\psi)_\phi\right)(C_1(t))\right] \\
&= (d\psi)_{c_1(t)}\left(\nabla_{g,\dot{c}_1} V_2\right)(c_1(t))
\end{aligned}$$

We say that the vector field $V(t) = V_2(c_1(t))$ is parallel along the curve $C_1(t) = C(t) = \psi(c(t))$, with $c(t) = c_1(t)$ if we have

$$\forall 1\leq k\leq p \qquad \frac{DV}{dt}(t) := \left(\nabla_{g,\dot{c}} V_2\right)(c(t)) = \dot{V}\ (t) + \sum_{1\leq i,j\leq p}\Gamma_{i,j}^k(c(t))\ \dot{c}^i\ (t)\ V^j(t) = 0 \tag{6.61}$$

Note that for any fixed initial vector field V' there always exists a vector field curve $V\ :\ t\in[0,1]\mapsto V(t)\in\mathbb{R}^p$ parallel to $c(t)$ s.t. $V(0) = V'$. In this case, one says that $V(1) = V''$ is obtained from $V(0) = V'$ by parallel transport along the curve c. Replacing $[0,1]$ by $[s,t]$, we obtain the following definition

$$\text{parall}_{c,s,t}\ :\ V'\in T_{c(s)}S_\phi \mapsto \text{parall}_{c,s,t}(V') = V(c(t))\in T_{c(s)}S_\phi \tag{6.62}$$

where $V(c(\tau))$, $\tau\in[s,t]$, is a unique vector field on $(c(\tau))$, $\tau\in[s,t]$, s.t.

$$V(c(s)) = V' \quad \text{and} \quad \left(\nabla_{g,\dot{c}} V_2\right)(c(\tau)) = 0$$

6.3.4 Second order covariant derivative

Tangent basis functions

We recall from (6.40) that

$$(\nabla\phi^i)_\psi = \sum_{1\leq k\leq p} g^{i,k}\ \partial_{\theta^k}\psi$$

Notice that

$$\partial_{\theta_m}\left((\nabla\phi^i)_\psi\right) = \sum_{1\leq k,l\leq p}\left(\partial_{\theta_m}(g^{i,k})\ \partial_{\theta^k}\psi + g^{i,k}\ \partial_{\theta_m,\theta_k}\psi\right) \tag{6.63}$$

Using the differential rule (6.4) we also prove that

$$\nabla^2\phi^i = \sum_{1\leq k\leq p}\left[\nabla\left(g_\phi^{i,k}\right)\ (\partial_{\theta^k}\psi)_\phi^T + g_\phi^{i,k}\ \nabla\left((\partial_{\theta^k}\psi)_\phi\right)\right]$$

Notice that

$$\partial_{x_l}\left(g_\phi^{i,k}\right) = \partial_{x_l}\left(g^{i,k}\circ\phi\right) = \sum_{1\leq m\leq p}\left(\partial_{\theta_m} g^{i,k}\right)_\phi\ \partial_{x_l}\phi^m \Rightarrow \partial\left(g_\phi^{i,k}\right) = \sum_{1\leq m\leq p}\left(\partial_{\theta_m} g^{i,k}\right)_\phi\ \partial\phi^m$$

from which we prove that

$$\nabla\left(g_\phi^{i,k}\right) \;=\; \pi\left(\partial\left(g_\phi^{i,k}\right)\right) = \sum_{1\leq m\leq p}\left(\partial_{\theta_m} g^{i,k}\right)_\phi \;\nabla\phi^m$$

On the other hand, we have

$$\partial_{x_m}\left(\left(\partial_{\theta_k}\psi^l\right)_\phi\right) = \textstyle\sum_{1\leq i\leq p}\left(\partial_{\theta_i,\theta_k}\psi^l\right)_\phi \;\partial_{x_m}\phi^i$$

$$\Rightarrow \partial\left(\left(\partial_{\theta_k}\psi^l\right)_\phi\right) = \textstyle\sum_{1\leq i\leq p}\left(\partial_{\theta_k,\theta_i}\psi^l\right)_\phi \;\partial\phi^i$$

$$\Rightarrow \nabla\left(\left(\partial_{\theta_k}\psi^l\right)_\phi\right) = \textstyle\sum_{1\leq i\leq p}\left(\partial_{\theta_k,\theta_i}\psi^l\right)_\phi \;\nabla\phi^i$$

This implies that

$$\partial\left(\left(\partial_{\theta_k}\psi\right)_\phi\right) = \left[\partial\left(\left(\partial_{\theta_k}\psi^1\right)_\phi\right),\ldots,\partial\left(\left(\partial_{\theta_k}\psi^r\right)_\phi\right)\right]$$

$$\Rightarrow \nabla\left(\left(\partial_{\theta_k}\psi\right)_\phi\right) = \left[\nabla\left(\left(\partial_{\theta_k}\psi^1\right)_\phi\right),\ldots,\nabla\left(\left(\partial_{\theta^k}\psi^r\right)_\phi\right)\right] = \textstyle\sum_{1\leq m\leq p}\;\nabla\phi^m\;\left(\partial_{\theta_k,\theta_m}\psi\right)_\phi^T$$

Using (6.63), we conclude that

$$\begin{aligned}\nabla^2\phi^i &= \sum_{1\leq m\leq p}\nabla\phi^m\;\partial_{\theta_m}\left(\left(\nabla\phi^i\right)_\psi^T\right) \\ &= \sum_{1\leq k,m\leq p}\left[\left(\partial_{\theta_m}g^{i,k}\right)_\phi\;\nabla\phi^m\left(\partial_{\theta_k}\psi\right)_\phi^T + g_\phi^{i,k}\;\nabla\phi^m\;\left(\partial_{\theta_k,\theta_m}\psi\right)_\phi^T\right]\end{aligned} \tag{6.64}$$

and

$$\begin{aligned}\mathrm{tr}\left(\nabla^2\phi^i\right) &= \sum_{1\leq k,m\leq p}\left[\left(\partial_{\theta_m}g^{i,k}\right)_\phi\;\left\langle\nabla\phi^m,\left(\partial_{\theta_k}\psi\right)_\phi\right\rangle + g_\phi^{i,k}\;\left\langle\nabla\phi^m,\left(\partial_{\theta_k,\theta_m}\psi\right)_\phi\right\rangle\right] \\ &= \sum_{1\leq m\leq p}\left[\left\langle\nabla\phi^m,\sum_{1\leq k\leq p}\left(\partial_{\theta_m}g^{i,k}\right)_\phi\left(\partial_{\theta_k}\psi\right)_\phi + \sum_{1\leq k\leq p}g_\phi^{i,k}\;\left(\partial_{\theta_k,\theta_m}\psi\right)_\phi\right\rangle\right]\end{aligned}$$

Using (6.63) this formula can be rewritten as follows

$$\mathrm{tr}\left(\nabla^2\phi^i\right)_\psi = \sum_{1\leq m\leq p}\left\langle\left(\nabla\phi^m\right)_\psi,\partial_{\theta_m}\left(\left(\nabla\phi^i\right)_\psi\right)\right\rangle$$

Using the fact that

$$\nabla\phi^m = \sum_{1\leq l\leq p}g_\phi^{m,l}\;\left(\partial_{\theta_l}\psi\right)_\phi \quad\Longrightarrow\quad \left\langle\nabla\phi^m,\left(\partial_{\theta_k}\psi\right)_\phi\right\rangle = 1_{m=k}$$

we also have the following formulae.

$$\begin{aligned}(\Delta\phi^i)_\psi &= \operatorname{tr}\left(\nabla^2\phi^i\right)_\psi \\ &= \sum_{1\le m\le p}\partial_{\theta_m} g^{i,m} + \sum_{1\le k\le p} g^{i,k} \sum_{1\le m,l\le p} g^{m,l}\ \langle \partial_{\theta^l}\psi, \partial_{\theta_k,\theta_m}\psi\rangle \\ &= \sum_{1\le j\le p}\partial_{\theta_j} g^{i,j} + \frac{1}{2}\sum_{1\le j\le p} g^{i,j} \sum_{1\le k,l\le p} g^{k,l}\ \partial_{\theta_j} g_{k,l} \\ &= \sum_{1\le j\le p}\frac{1}{\sqrt{\det(g)}}\,\partial_{\theta_j}\left(\sqrt{\det(g)}\ g^{i,j}\right) = \operatorname{div}\left(\nabla\phi^i\right)_\psi \end{aligned} \tag{6.65}$$

The last assertion is a direct consequence of (6.54) applied to the vector field

$$W = \nabla\phi^i = \sum_{1\le j\le p} g_\phi^{i,j}\ \left(\partial_{\theta_j}\psi\right)_\phi \tag{6.66}$$

We end this section with a formula relating $\Delta\phi^i$ to the Christoffel symbols introduced in (6.58). Firstly, we observe that

$$\sum_{1\le m\le m}\partial_{\theta_p}\left\langle(\nabla\phi^m)_\psi, (\nabla\phi^i)_\psi\right\rangle = \sum_{1\le m\le p}\partial_{\theta_m} g^{i,m} \tag{6.67}$$

On the other hand, we have

$$\begin{aligned}&\textstyle\sum_{1\le m\le p}\partial_{\theta_m}\left\langle(\nabla\phi^m)_\psi, (\nabla\phi^i)_\psi\right\rangle \\ &= \textstyle\sum_{1\le m\le p}\left\langle\partial_{\theta_m}\left((\nabla\phi^m)_\psi\right), (\nabla\phi^i)_\psi\right\rangle + \sum_{1\le m\le p}\left\langle(\nabla\phi^m)_\psi, \partial_{\theta_m}\left((\nabla\phi^i)_\psi\right)\right\rangle \\ &= \textstyle\sum_{1\le m,l\le p}\left\langle\partial_{\theta_m}\left(g^{m,l}\ \partial_{\theta_l}\psi\right), (\nabla\phi^i)_\psi\right\rangle + \left(\Delta\phi^i\right)_\psi \\ &= \textstyle\sum_{1\le m,l\le p}\partial_{\theta_m} g^{m,l}\ \left\langle\ \partial_{\theta^l}\psi, (\nabla\phi^i)_\psi\right\rangle + \sum_{1\le m,l\le p} g^{m,l}\ \left\langle\partial_{\theta_m\theta_l}\psi, (\nabla\phi^i)_\psi\right\rangle + \left(\Delta\phi^i\right)_\psi\end{aligned}$$

Combined with (6.67), this implies that

$$\sum_{1\le m\le p}\partial_{\theta_m} g^{i,m} = \sum_{1\le m\le p}\partial_{\theta_m} g^{m,i} + \sum_{1\le m,l\le p} g^{m,l}\ \underbrace{\left\langle\partial_{\theta_m\theta_l}\psi, (\nabla\phi^i)_\psi\right\rangle}_{=\Gamma^i_{m,l}} + \left(\Delta\phi^i\right)_\psi$$

Using (6.65), we conclude that

$$\left(\Delta\phi^i\right)_\psi = -\sum_{1\le m,l\le p} g^{m,l}\ \Gamma^i_{m,l} = \operatorname{div}\left(\nabla\phi^i\right)_\psi = \sum_{1\le j\le p}\frac{1}{\sqrt{\det(g)}}\,\partial_{\theta_j}\left(\sqrt{\det(g)}\ g^{i,j}\right) \tag{6.68}$$

Composition formulae

Suppose we are given a function $F = f\circ\phi$ on S. By (6.48) we have

$$\nabla F = \nabla(f\circ\phi) = \sum_{1\le j\le p}\left(\partial_{\theta_j} f\right)_\phi\ \nabla\phi^j \tag{6.69}$$

and

$$F = \left(\partial_{\theta_j} f\right)_\phi = (\partial_{\theta_j} f) \circ \phi \Longrightarrow \nabla\left(\left(\partial_{\theta_j} f\right)_\phi\right) = \sum_{1\leq i\leq p} \left(\partial_{\theta_i,\theta_j} f\right)_\phi \ \nabla\phi^i$$

Using the differential rule (6.4) we find that

$$\nabla^2 F \ = \ \sum_{1\leq j\leq p} \left[\nabla\left(\left(\partial_{\theta_j} f\right)_\phi\right) \ \left(\nabla\phi^j\right)^T + \left(\partial_{\theta_j} f\right)_\phi \ \nabla^2\phi^j\right]$$

By (6.64), this yields the second covariant derivative formula

$$\begin{aligned}\nabla^2(f\circ\phi) \ &= \ \sum_{1\leq i,j\leq p} \left(\partial_{\theta_i,\theta_j} f\right)_\phi \ \nabla\phi^i \ \left(\nabla\phi^j\right)^T + \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right)_\phi \ \nabla^2\phi^j \\ &= \ \sum_{1\leq i,j\leq p} \left[\left(\partial_{\theta_i,\theta_j} f\right)_\phi \ \nabla\phi^i \ \left(\nabla\phi^j\right)^T + \left(\partial_{\theta_j} f\right)_\phi \ \nabla\phi^i \ \left[\partial_{\theta_i}\left(\left(\nabla\phi^j\right)^T_\psi\right)\right]_\phi\right] \end{aligned} \tag{6.70}$$

We also readily check that

$$\begin{aligned}\mathrm{tr}\left(\nabla^2 F\right) \ &= \ \sum_{1\leq i,j\leq p} \left(\partial_{\theta_i,\theta_j} f\right)_\phi \ \left\langle\nabla\phi^i, \nabla\phi^j\right\rangle + \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right)_\phi \ \mathrm{tr}\left(\nabla^2\phi^j\right) \\ &= \ \sum_{1\leq i,j\leq p} g^{i,j}_\phi \ \left(\partial_{\theta_i,\theta_j} f\right)_\phi \ + \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right)_\phi \ \mathrm{tr}\left(\nabla^2\phi^j\right)\end{aligned}$$

from which we find the Laplacian formula

$$\Delta(f\circ\phi) \ := \ \mathrm{tr}\left(\nabla^2(f\circ\phi)\right) = \sum_{1\leq i,j\leq p} g^{i,j}_\phi \ \left(\partial_{\theta_i,\theta_j} f\right)_\phi \ + \sum_{1\leq j\leq p} \left(\partial_{\theta_j} f\right)_\phi \ \Delta\phi^j \tag{6.71}$$

Using (6.68) we also have

$$\Delta(f\circ\phi) \ := \ \mathrm{tr}\left(\nabla^2(f\circ\phi)\right) = \sum_{1\leq l,m\leq p} g^{l,m}_\phi \left[\ \left(\partial_{\theta_l,\theta_m} f\right)_\phi - \sum_{1\leq j\leq p} \Gamma^j_{\phi,l,m} \ \left(\partial_{\theta_j} f\right)_\phi\right]$$

with $\Gamma^j_{\phi,l,m} = \Gamma^j_{l,m}\circ\phi$.

Using (6.68), we have a divergence formulation of the Riemannian Laplacian:

$$\begin{aligned}\left(\Delta(f\circ\phi)\right)_\psi \ &= \ \sum_{1\leq i\leq p}\left[\sum_{1\leq j\leq p} g^{i,j}\partial_{\theta_i}(\partial_{\theta_j} f) + \frac{1}{\sqrt{\det(g)}}\ \partial_{\theta_i}\left(\sqrt{\det(g)}\ \sum_{1\leq j\leq p} g^{i,j}\right)\ \partial_{\theta_j} f\right] \\ &= \ \sum_{1\leq i\leq p} \frac{1}{\sqrt{\det(g)}}\ \partial_{\theta_i}\left(\sqrt{\det(g)}\ \sum_{1\leq j\leq p} g^{i,j}\ \partial_{\theta_j} f\right) := \mathrm{div}_g\left(\nabla_g(f)\right) := \Delta_g(f)\end{aligned} \tag{6.72}$$

For any couple of functions f_1 and f_2 we also quote the following formula

$$\mathrm{div}_g\left(f_1\ \nabla_g(f_2)\right) = f_1\ \mathrm{div}_g\left(\nabla_g(f_2)\right) + \left\langle\nabla_g f_1, \nabla_g f_2\right\rangle_g \tag{6.73}$$

We check this claim using the fact that

$$\partial_{\theta_i}\left(\sqrt{\det(g)}\ f_1\ \sum_{1\leq j\leq p} g^{i,j}\ \partial_{\theta_j} f_2\right)$$

$$= f_1\ \partial_{\theta_i}\left(\sqrt{\det(g)}\ \sum_{1\leq j\leq p} g^{i,j}\ \partial_{\theta_j} f_2\right) + \partial_{\theta_i}(f_1) \times \sqrt{\det(g)}\ \ \sum_{1\leq j\leq p} g^{i,j}\ \partial_{\theta_j} f_2$$

and

$$\sum_{1\leq i\leq p} \partial_{\theta_i}(f_1) \sum_{1\leq i,j\leq p} g^{i,j}\ \partial_{\theta_j} f_2 = \sum_{1\leq i,j\leq p} g^{i,j}\ \partial_{\theta_i} f_1\ \partial_{\theta_j} f_2 = \langle \partial f_1, \partial f_2\rangle_{g^{-1}} = \langle \nabla_g f_1, \nabla_g f_2\rangle_g$$

Hessian operators

We end this section with an Hessian interpretation of the second covariant derivative $\nabla^2 F$. We let W_1, W_2 be a couple of vector fields in $T(S)$. We let V_1, V_2 their pull back vector fields so that

$$W_k \circ \psi = (d\psi)(V_k) := \sum_{1\leq m\leq p} V_k^m\ \partial_{\theta_m}\psi$$

for any $k = 1, 2$. Using (6.70) we prove that

$$\begin{aligned} W_1^T\ \nabla^2 F\ W_2 \ &= \sum_{1\leq m,m'\leq p} V_{\phi,1}^m V_{\phi,2}^{m'} \sum_{1\leq i,j\leq p} \left(\partial_{\theta_i,\theta_j} f\right)_\phi\ \overbrace{\left(\partial_{\theta_m}\psi\right)_\phi^T \nabla\phi^i}^{=1_{m=i}}\ \overbrace{\left(\nabla\phi^j\right)^T \left(\partial_{\theta_{m'}}\psi\right)_\phi}^{=1_{m'=j}} \\ &+ \sum_{1\leq m,m'\leq p} V_{\phi,1}^m V_{\phi,2}^{m'} \sum_{1\leq i,j\leq p} \left(\partial_{\theta_j} f\right)_\phi\ \underbrace{\left(\partial_{\theta_m}\psi\right)_\phi^T \nabla\phi^i}_{=1_{m=i}}\ \underbrace{\left[\partial_{\theta_i}\left(\left(\nabla\phi^j\right)_\psi^T\right)\right]_\phi \left(\partial_{\theta_{m'}}\psi\right)_\phi}_{=-\Gamma^j_{\phi,i,m'}} \end{aligned}$$

The r.h.s. assertion comes from the fact that

$$\partial_{\theta_i}\left(\underbrace{\left(\nabla\phi^j\right)_\psi^T\left(\partial_{\theta_m}\psi\right)}_{=1_{m=j}}\right) = 0 \Rightarrow \partial_{\theta_i}\left(\left(\nabla\phi^j\right)_\psi^T\right)\left(\ \partial_{\theta_m}\psi\right) = -\left(\nabla\phi^j\right)_\psi^T\ \ \left(\partial_{\theta_i,\theta_m}\psi\right) = -\Gamma^j_{i,m}$$

This yields

$$W_1^T\nabla^2 F\ W_2 \ = \sum_{1\leq m,m'\leq p} V_{\phi,1}^m V_{\phi,2}^{m'}\ \left[\left(\partial_{\theta_m,\theta_{m'}} f\right)_\phi - \sum_{1\leq j\leq p} \Gamma^j_{\phi,m,m'}\ \left(\partial_{\theta_j} f\right)_\phi\ \right] = V_{\phi,1}^T\ \left(\mathrm{Hess}_g(f)\right)_\phi\ V_{\phi,2}$$

or equivalently

$$\left(W_1^T\nabla^2 F\ W_2\right)_\psi = \langle W_1, \nabla^2 F\ W_2\rangle = V_1^T\ \mathrm{Hess}_g(f)\ V_2 = \langle V_1, \mathrm{Hess}_g(f)\ V_2\rangle = \langle V_1, \nabla_g^2 f\ V_2\rangle_g$$

with the Hessian matrix field $\mathrm{Hess}_g(f) = \left((\mathrm{Hess}_g(f))_{m,m'}\right)_{1\leq m,m'\leq p}$ on S_ψ with entries

$$(\mathrm{Hess}_g(f))_{m,m'} = \partial_{\theta_m,\theta_{m'}} f - \sum_{1\leq j\leq p} \Gamma^j_{m,m'}\ \partial_{\theta_j} f$$

and the Riemannian second covariant derivative

$$\nabla_g^2 = g^{-1}\mathrm{Hess}_g(f) \quad \left(\Rightarrow \langle V_1, \nabla_g^2 f\ V_2\rangle_g = V_1^T\ g\ \left(g^{-1}\mathrm{Hess}_g(f)\right)\ V_2 = V_1^T\ \mathrm{Hess}_g(f)\ V_2\right)$$

In differential calculus literature, the above formulae are sometimes written in the following form

$$\left(\nabla^2 F\right)\ (W_1,,W_2) := W_1^T \nabla^2 F\ W_2 = V_{\phi,1}^T\ \left(\nabla_g^2 f\right)_\phi\ V_{\phi,2} := \left(\left(\nabla_g^2 f\right)(V_1,V_2)\right)\circ\phi$$

or using the Hessian symbol

$$\mathrm{Hess}(F)\ (W_1,W_2) := \left(\nabla^2 F\right)\ (W_1,,W_2) = \left(\left(\nabla_g^2 f\right)(V_1,V_2)\right)\circ\phi := \left(\mathrm{Hess}_g(f)(V_1,V_2)\right)\circ\phi$$

In view of (6.57) and (6.59) we also have

$$\begin{aligned}
\left(\partial_{W_1}(\partial_{W_2}(F))\right) &= \sum_{1\leq m,m'\leq p} V_{\phi,1}^m\ \left(\partial_{\theta_m} V_2^{m'}\right)_\phi\ \left(\partial_{\theta_{m'}} f\right)_\phi + \sum_{1\leq m,m'\leq p} V_{\phi,1}^m V_{\phi,2}^{m'}\ \left(\partial_{\theta_m,\theta'_m} f\right)_\phi \\
\nabla_{W_1} W_2 &= \sum_{1\leq m'\leq p}\left[\sum_{1\leq m\leq p} V_{\phi,1}^m\ \left(\partial_{\theta_m} V_2^{m'}\right)_\phi + \sum_{1\leq m,j\leq p}\Gamma_{\phi,m,j}^{m'}\ V_{\phi,1}^m\ V_{\phi,2}^j\right]\ \left(\partial_{\theta_{m'}}\psi\right)_\phi
\end{aligned}$$

Since

$$\left(\partial_{\theta_k}\psi\right)^T (\partial F)_\psi = \sum_{0\leq l\leq r}\ \left(\partial_{x_l} F\right)_\psi\ \partial_{\theta_k}\psi^l = \partial_{\theta_k}\left(F\circ\psi\right) = \partial_{\theta_k} f$$

we find that

$$\left(\nabla_{W_1} W_2\right)^T \partial F = \sum_{1\leq m,m'\leq p} V_{\phi,1}^m\ \left(\partial_{\theta_m} V_2^{m'}\right)_\phi \left(\partial_{\theta_{m'}} f\right)_\phi + \sum_{1\leq m,m'\leq p} V_{\phi,1}^m\ V_{\phi,2}^{m'} \sum_{1\leq j\leq p}\Gamma_{\phi,m,m'}^j\ \left(\partial_{\theta_j} f\right)_\phi$$

and therefore

$$W_1^T\ \nabla^2 F\ W_2 = \left(\partial_{W_1}(\partial_{W_2}(F))\right) - \left(\nabla_{W_1} W_2\right)^T \partial F$$

Last but not least, for any vector fields $W_1, W_2, W_3 \in T(S)$ we have

$$\partial_{W_1}\left(\langle W_2, W_3\rangle\right) = \sum_{1\leq k,l\leq r} W_1^k\ \left[W_2^l \partial_{x_k}\left(W_3^l\right) + \partial_{x_k}\left(W_2^l\right)\ W_3^l\right] = \langle \partial_{W_1} W_2, W_3\rangle + \langle W_2, \partial_{W_1} W_3\rangle$$

Notice that

$$W_3 \in T(S) \implies \langle \partial_{W_1} W_2, W_3\rangle = \langle \pi(\partial_{W_1} W_2), W_3\rangle = \langle \nabla_{W_1} W_2, W_3\rangle$$

so that

$$\partial_{W_1}\left(\langle W_2, W_3\rangle\right) = \langle \nabla_{W_1} W_2, W_3\rangle + \langle W_2, \nabla_{W_1} W_3\rangle \tag{6.74}$$

On the other hand, we have

$$\partial_{W_2}(F) = \langle W_2, \partial F\rangle = \langle W_2, \nabla F\rangle$$

$$\Rightarrow \partial_{W_1}(\partial_{W_2}(F)) = \partial_{W_1}\langle W_2, \nabla F\rangle = \langle \nabla_{W_1} W_2, \nabla F\rangle + \langle W_2, \nabla_{W_1}\nabla F\rangle$$

and therefore

$$\begin{aligned}
W_1^T\ \nabla^2 F\ W_2 &= \left(\partial_{W_1}(\partial_{W_2}(F))\right) - \left(\nabla_{W_1} W_2\right)^T \partial F \\
&= \left(\partial_{W_1}(\partial_{W_2}(F))\right) - \langle \nabla_{W_1} W_2, \partial F\rangle \\
&= \left(\partial_{W_1}(\partial_{W_2}(F))\right) - \langle \nabla_{W_1} W_2, \nabla F\rangle = \langle W_2, \nabla_{W_1}\nabla F\rangle
\end{aligned}$$

We end this section with some comments on the parallel transport technique introduced in (6.61). By (6.56), for and $F = f\circ\phi$ we have

$$\partial_{W_1}(F)\circ\psi = \langle \nabla_g f, V_1\rangle_g = \partial_{V_1}(f)$$

On the other hand, we have

$$\langle W_2, W_3\rangle = \langle V_2, V_3\rangle_g \circ \phi \Rightarrow (\partial_{W_1} \langle W_2, W_3\rangle) \circ \psi = \partial_{V_1} \langle V_2, V_3\rangle_g$$

and by (6.60) we prove that

$$\langle \nabla_{W_1} W_2, W_3\rangle \circ \psi = \langle (d\psi)(\nabla_{g,V_1} V_2), (d\psi)(V_3)\rangle = \langle \nabla_{g,V_1} V_2, V_3\rangle_g$$

Combining these results with (6.74) we conclude that

$$\partial_{V_1} \langle V_2, V_3\rangle_g = \langle \nabla_{g,V_1} V_2, V_3\rangle_g + \langle V_2, \nabla_{g,V_1} V_3\rangle_g$$

We let $c_1(t)$ be a given curve in S_ϕ with $\dot{c}_1(t) = V_1(c_1(t))$, and $U_i : t \in [0,1] \mapsto U_i(t) = V_i(c_1(t)) \in \mathbb{R}^p$ two parallel vectors to $c(t)$ s.t. $U_i(0) = V_i(c_1(0))$, with $i = 2, 3$. In this situation, using (6.61) we have

$$\begin{aligned} \frac{d}{dt} \langle V_2(c_1(t)), V_3(c_1(t))\rangle_{g(c_1(t))} &= \left(\partial_{V_1} \left(\langle V_2, V_3\rangle_g\right)\right)(c_1(t)) \\ &= \left(\langle \nabla_{g,V_1} V_2, V_3\rangle_g + \langle V_2, \nabla_{g,V_1} V_3\rangle_g\right)(c_1(t)) = 0 \end{aligned}$$

This shows that the parallel transport is an isometry

$$\langle V_2(c_1(0)), V_3(c_1(0))\rangle_{g(c_1(0))} = \langle V_2(c_1(1)), V_3(c_1(1))\rangle_{g(c_1(1))} \tag{6.75}$$

6.4 Stochastic calculus in chart spaces

6.4.1 Brownian motion in Riemannian manifolds

We let $\Theta_t = \begin{pmatrix} \Theta_t^1 \\ \vdots \\ \Theta_t^p \end{pmatrix}$ be the $\mathbb{R}^p$-diffusion on the parameter space S_ϕ defined by

$$\begin{aligned} \forall 1 \leq i \leq p \qquad d\Theta_t^i &= \frac{1}{2} \left(\Delta \phi^i\right)_\psi (\Theta_t)\, dt + \left(\nabla \phi^i\right)_\psi^T (\Theta_t)\, dB_t \\ &= -\sum_{1 \leq j,k \leq p} g^{j,k}(\Theta_t)\, \Gamma_{j,k}^i(\Theta_t)\, dt + \left(\nabla \phi^i\right)_\psi^T (\Theta_t)\, dB_t \quad (\Leftarrow (6.68)) \end{aligned} \tag{6.76}$$

where B_t stands for a standard r-dimensional Brownian motion.

Notice that

$$
\begin{aligned}
d\Theta_t d\Theta_t^T &= \begin{pmatrix} \left(\nabla\phi^1\right)_\psi^T \\ \vdots \\ \left(\nabla\phi^p\right)_\psi^T \end{pmatrix}(\Theta_t)\; dB_t dB_t^T \left(\left(\nabla\phi^1\right)_\psi, \ldots, \left(\nabla\phi^p\right)_\psi\right)(\Theta_t) \\
&= \begin{pmatrix} \left(\nabla\phi^1\right)_\psi^T \left(\nabla\phi^1\right)_\psi & \ldots & \left(\nabla\phi^1\right)_\psi^T \left(\nabla\phi^r\right)_\psi \\ \vdots & & \vdots \\ \left(\nabla\phi^r\right)_\psi^T \left(\nabla\phi^1\right)_\psi & \ldots & \left(\nabla\phi^r\right)_\psi^T \left(\nabla\phi^r\right)_\psi \end{pmatrix}(\Theta_t) \\
&= \begin{pmatrix} \left\langle \left(\nabla\phi^1\right)_\psi, \left(\nabla\phi^1\right)_\psi \right\rangle & \ldots & \left\langle \left(\nabla\phi^1\right)_\psi, \left(\nabla\phi^r\right)_\psi \right\rangle \\ \vdots & & \vdots \\ \left\langle \left(\nabla\phi^r\right)_\psi, \left(\nabla\phi^1\right)_\psi \right\rangle & \ldots & \left\langle \left(\nabla\phi^r\right)_\psi, \left(\nabla\phi^r\right)_\psi \right\rangle \end{pmatrix}(\Theta_t)
\end{aligned}
$$

Thus, using Ito formula for any smooth function f on $\mathbb{R}^p$ we have

$$
\begin{aligned}
df(\Theta_t) &= \sum_{1\le i\le p} \partial_{\theta_i}(f)(\Theta_t)\; d\Theta_t^i + \frac{1}{2}\sum_{1\le i,j\le p} \partial_{\theta_i,\theta_j}(f)(\Theta_t)\; d\Theta_t^i d\Theta_t^j \\
&= \mathcal{L}(f)(\Theta_t)dt + d\mathcal{M}_t(f)
\end{aligned}
$$

with the infinitesimal generator $\mathcal{L}$ associated with the diffusion process Θ_t is given by

$$
\begin{aligned}
\mathcal{L}(f) &= \frac{1}{2}\left[\sum_{1\le i\le p} \partial_{\theta_i} f\; \left(\Delta\phi^i\right)_\psi + \sum_{1\le i,j\le p} \partial_{\theta_i,\theta_j}(f)\; \left\langle \left(\nabla\phi^i\right)_\psi, \left(\nabla\phi^j\right)_\psi \right\rangle\right] \\
&= \frac{1}{2}\left[\sum_{1\le i\le p} \partial_{\theta_i} f\; \left(\Delta\phi^i\right)_\psi + \sum_{1\le i,j\le p} g^{i,j}\; \partial_{\theta_i,\theta_j} f\right] \\
&= \frac{1}{2}\,\Delta_g(f) = \frac{1}{2}\,\mathrm{div}_g\left(\nabla_g(f)\right) \quad (\Leftarrow (6.72))
\end{aligned} \tag{6.77}
$$

and the martingale $\mathcal{M}_t(f)$ defined by

$$
d\mathcal{M}_t(f) = \sum_{1\le i\le p} \partial_{\theta_i}(f)(\Theta_t)\; \left(\nabla\phi^i\right)_\psi^T(\Theta_t)\; dB_t
$$

Using (6.69) and (6.71) we find that

$$
\begin{aligned}
\sum_{1\le i\le p} \left(\partial_{\theta_i} f\right)_\phi\; \left(\nabla\phi^i\right)^T &= \nabla(f\circ\phi) = \nabla F \\
\mathcal{L}(f)\circ\phi &= \frac{1}{2}\,\Delta(f\circ\phi) = \Delta(F) \quad \text{with} \quad F = f\circ\phi
\end{aligned}
$$

Therefore, if we set $X_t = \psi(\Theta_t) \;\Rightarrow \Theta_t = \phi(X_t)$ we find that

$$
df(\Theta_t) = d(f\circ\phi)(X_t) = dF(X_t) = \frac{1}{2}\,\Delta(F)(X_t)dt + dM_t(F)
$$

with the martingale

$$
dM_t(F) = (\nabla F)^T(X_t)\; dB_t
$$

Choosing $f = \psi^k \Rightarrow F = \psi^k \circ \phi = \chi^k$ (cf. (6.51)) we find that

$$\pi(X_t)\ dB_t = \begin{bmatrix} (\nabla\ \chi^1)^T \\ \vdots \\ (\nabla\ \chi^r)^T \end{bmatrix} (X_t)\ dB_t$$

Combining this observation with (6.53) we find that

$$\begin{aligned} dX_t^k &= d\psi^k(\Theta_t) = \frac{1}{2}\ \Delta(\chi^k)(X_t)\ dt + (\nabla\ \chi^k)^T(X_t)\ dB_t \\ &= -\frac{1}{2}\ \mathbb{H}^k(X_t)\ dt + \sum_{1\le j\le r} \pi_j^k(X_t)\ dB_t^j \quad (\Leftarrow (6.52)) \end{aligned} \tag{6.78}$$

6.4.2 Diffusions in chart spaces

Starting from the equation (6.78), if we set $\Theta_t^i = \phi^i(X_t)$ then we find that

$$\begin{aligned} (6.47) \quad &\Rightarrow \quad \nabla\ \chi^k = \nabla(\psi^k \circ \phi) = \textstyle\sum_{1\le j\le p} \left(\partial_{\theta_j}\psi^k\right)_\phi\ \nabla\phi^j \\ &\Rightarrow \quad (\nabla\ \chi^k)^T(X_t)\ dB_t = \textstyle\sum_{1\le j\le p} \left(\partial_{\theta_j}\psi^k\right)_\phi\ (\nabla\phi^j)^T(X_t)\ dB_t \\ &\qquad\qquad = \textstyle\sum_{1\le j\le p} \left(\partial_{\theta_j}\psi^k\right)_\phi\ \left\langle \nabla\phi^j(X_t)\ dB_t \right\rangle \end{aligned}$$

and

$$\begin{aligned} dX_t^k dX_t^l &= \sum_{1\le i,j\le p} \left(\partial_{\theta_i}\psi^k\right)_\phi\ \left(\partial_{\theta_j}\psi^l\right)_\phi\ \left\langle \nabla\phi^i(X_t)\ dB_t \right\rangle \left\langle \nabla\phi^j(X_t)\ dB_t \right\rangle \\ &= \sum_{1\le i,j\le p} \left(\partial_{\theta_i}\psi^k\right)_\phi\ \left(\partial_{\theta_j}\psi^l\right)_\phi\ \left\langle \nabla\phi^i(X_t), \nabla\phi^j(X_t) \right\rangle\ dt \end{aligned}$$

Therefore, using Ito formula we have

$$\begin{aligned} d\phi^i(X_t) &= \sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right)(X_t)\ \left[\frac{1}{2}\ \Delta(\chi^k)(X_t)\ dt + (\nabla\ \chi^k)^T(X_t)\ dB_t\right] \\ &\quad + \frac{1}{2} \sum_{1\le k,l\le r} \left(\partial_{x_l,x_k}\phi^i\right)(X_t) \sum_{1\le i,j\le p} \left(\partial_{\theta_i}\psi^k\right)_\phi\ \left(\partial_{\theta_j}\psi^l\right)_\phi\ \left\langle \nabla\phi^i(X_t), \nabla\phi^j(X_t) \right\rangle\ dt \end{aligned}$$

Notice that

$$\begin{aligned} \sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right)\ (\nabla\ \chi^k)^T &= \sum_{1\le j\le p} \left[\sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right) \left(\partial_{\theta_j}\psi^k\right)_\phi\right]\ \left(\nabla\phi^j\right)^T \\ &= \sum_{1\le j\le p} \left(\partial_{\theta_j}\left(\phi^i \circ \psi\right)\right)_\phi \left(\nabla\phi^j\right)^T = \left(\nabla\phi^i\right)^T \quad \left(\Leftarrow (\phi^i \circ \psi)(v) = v^i\right) \end{aligned} \tag{6.79}$$

and using (6.71) we have

$$\begin{aligned}
&\Delta(\chi^k) = \Delta(\psi^k \circ \phi) = \textstyle\sum_{1\le i,j\le p} \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle \ \left(\partial_{\theta_i,\theta_j}\psi^k\right)_\phi \ + \sum_{1\le j\le p} \left(\partial_{\theta_j}\psi^k\right)_\phi \ \Delta\phi^j \\
&\Rightarrow \sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right) \Delta(\chi^k) + \sum_{1\le k,l\le r} \left(\partial_{x_l,x_k}\phi^i\right) \ \sum_{1\le i,j\le p} \left(\partial_{\theta_i}\psi^k\right)_\phi \ \left(\partial_{\theta_j}\psi^l\right)_\phi \ \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle \\
&= \sum_{1\le i,j\le p} \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle \ \left[\sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right) \left(\partial_{\theta_i,\theta_j}\psi^k\right)_\phi + \sum_{1\le k,l\le r} \left(\partial_{x_l,x_k}\phi^i\right) \ \left(\partial_{\theta_i}\psi^k\right)_\phi \ \left(\partial_{\theta_j}\psi^l\right)_\phi \right] \\
&\qquad\qquad + \sum_{1\le j\le p} \sum_{1\le k\le r} \left(\partial_{x_k}\phi^i\right) \left(\partial_{\theta_j}\psi^k\right)_\phi \ \Delta\phi^j \\
&= \sum_{1\le i,j\le p} \left\langle \nabla\phi^i, \nabla\phi^j \right\rangle \ \left(\partial_{\theta_i,\theta_j}(\phi^i \circ \psi)\right)_\phi + \sum_{1\le j\le p} \left(\partial_{\theta_j}(\phi^i \circ \psi)\right)_\phi \ \Delta\phi^j = \ \Delta\phi^i
\end{aligned}$$

This implies that

$$\forall 1 \le i \le p \qquad d\phi^i(X_t) = \frac{1}{2} \ \left(\Delta\phi^i\right)(X_t)dt + \left(\nabla\phi^i\right)^T (X_t)dB_t \tag{6.80}$$

Letting $\Theta_t := \phi(X_t) \ \Rightarrow X_t = \psi(\Theta_t)$ we arrive at the equation

$$d\Theta_t = \frac{1}{2} \ \left(\Delta\phi^i\right)_\psi (\Theta_t) \ dt + \left(\nabla\phi^i\right)_\psi^T (\Theta_t) \ dB_t$$

6.5 Somme illustrations

6.5.1 Brownian motion on spheres

The unit circle $S = \mathbb{S}^1 \subset \mathbb{R}^2$

The unit circle can be described in terms of the polar coordinates mapping

$$\psi(\theta) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

In this situation, we can check (cf. (6.84) and (6.85)) that

$$(\nabla\phi)_\psi = \partial_\theta\psi = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} \quad \text{and} \quad (\Delta\phi)_\psi = 0$$

In this situation, we have

$$(6.76) \quad \Longleftrightarrow \quad d\Theta_t = (\nabla\phi)_\psi^T(\Theta_t) \ dB_t = -\sin(\Theta_t) \ dB_t^1 + \cos(\Theta_t) \ dB_t^1 := d\overline{B}_t$$

Notice that $\overline{B}_t$ is itself a standard Brownian motion

$$d\overline{B}_t d\overline{B}_t = \left(\cos^2(\Theta_t) + \sin^2(\Theta_t)\right) \ dt = dt$$

The unit sphere $S = \mathbb{S}^2 \subset \mathbb{R}^3$

The 2-sphere can be parametrized by the spherical coordinates mapping

$$\psi(\theta) = \begin{pmatrix} \sin(\theta_1)\cos(\theta_2) \\ \sin(\theta_1)\sin(\theta_2) \\ \cos(\theta_1) \end{pmatrix}$$

In this situation, we can check (cf. (6.86) and (6.87))

$$\begin{array}{rcll} (\nabla\phi^1)_\psi(\theta) & = & \begin{pmatrix} \cos(\theta_1)\cos(\theta_2) \\ \cos(\theta_1)\sin(\theta_2) \\ -\sin(\theta_1) \end{pmatrix} & (\nabla\phi^2)_\psi(\theta) = \dfrac{1}{\sin(\theta_1)} \begin{pmatrix} -\sin(\theta_2) \\ \cos(\theta_2) \\ 0 \end{pmatrix} \\ (\Delta\phi^1)_\psi(\theta) & = & \cot(\theta_1) & (\Delta\phi^2)_\psi(\theta) = 0 \end{array}$$

In this situation, we have

(6.76)

$$\iff \left\{ \begin{array}{rcl} d\Theta_t^1 & = & \dfrac{1}{2}\cot(\Theta_t^1)\,dt + \left[\cos(\Theta_t^1)\left(\cos(\Theta_t^2)dB_t^1 + \sin(\Theta_t^2)dB_t^2\right) - \sin(\Theta_t^1)\,dB_t^3\right] \\ & := & \dfrac{1}{2}\cot(\Theta_t^1)\,dt + d\overline{B}_t^1 \\ d\Theta_t^2 & = & \dfrac{1}{\sin(\Theta_t^1)}\left[-\sin(\Theta_t^2)\,dB_t^1 + \cos(\Theta_t^2)\,dB_t^2\right] := \dfrac{1}{\sin(\Theta_t^1)}\,d\overline{B}_t^2 \end{array} \right.$$

Notice that

$$d\overline{B}_t^1 d\overline{B}_t^2 = 0 \quad \text{and} \quad d\overline{B}_t^1 d\overline{B}_t^1 = dt = d\overline{B}_t^2 d\overline{B}_t^2$$

so that (6.76) can be rewritten as follows

$$\left\{ \begin{array}{rcl} d\Theta_t^1 & = & \dfrac{1}{2}\cot(\Theta_t^1)\,dt + dB_t^1 \\ d\Theta_t^2 & = & \dfrac{1}{\sin(\Theta_t^1)}\,dB_t^2 \end{array} \right.$$

6.5.2 Brownian motion on the Torus

The 2-Torus is the null level set of the function

$$\varphi(x) = \left(R - \sqrt{x_1^2 + x_2^2}\right)^2 + x_3^2 - r^2$$

with $r < R$. It can be parametrized by the spherical coordinates mapping

$$\psi(\theta) = \begin{pmatrix} (R + r\cos(\theta_1))\cos(\theta_2) \\ (R + r\cos(\theta_1))\sin(\theta_2) \\ r\sin(\theta_1) \end{pmatrix}$$

In this situation, we can check (cf. (6.86) and (6.87))

$$\begin{array}{rcll} (\nabla\phi^1)_\psi(\theta) & = & r^{-1}\begin{pmatrix} -\sin(\theta_1)\cos(\theta_2) \\ -\sin(\theta_1)\sin(\theta_2) \\ \cos(\theta_1) \end{pmatrix} & (\nabla\phi^2)_\psi(\theta) = (R + r\cos(\theta_1))^{-1}\begin{pmatrix} -\sin(\theta_2) \\ \cos(\theta_2) \\ 0 \end{pmatrix} \\ (\Delta\phi^1)_\psi(\theta) & = & -\dfrac{\sin(\theta_1)}{r(R + r\cos(\theta_1))} & (\Delta\phi^2)_\psi(\theta) = 0 \end{array}$$

In this situation, we have

$$(6.76) \iff \begin{cases} d\Theta_t^1 &= -\dfrac{\sin(\Theta_t^1)}{2r(R+r\cos(\Theta_t^1))}\,dt \\ & \quad +\dfrac{1}{r}\left[-\sin\left(\Theta_t^1\right)\left(\cos\left(\Theta_t^2\right)\,dB_t^1+\sin\left(\Theta_t^2\right)\,dB_t^2\right)+\cos\left(\Theta_t^1\right)\,dB_t^3\right] \\ d\Theta_t^2 &= \dfrac{1}{(R+r\cos(\Theta_t^1))}\,\left[-\sin\left(\Theta_t^2\right)\,dB_t^1+\cos\left(\Theta_t^2\right)\,dB_t^2\right] \end{cases}$$

We set

$$\begin{aligned} d\overline{B}_t^1 &= -\sin\left(\Theta_t^1\right)\left(\cos\left(\Theta_t^2\right)\,dB_t^1+\sin\left(\Theta_t^2\right)\,dB_t^2\right)+\cos\left(\Theta_t^1\right)\,dB_t^3 \\ d\overline{B}_t^2 &= -\sin\left(\Theta_t^2\right)\,dB_t^1+\cos\left(\Theta_t^2\right)\,dB_t^2 \end{aligned}$$

It is readily checked that

$$d\overline{B}_t^1 d\overline{B}_t^2 = 0 \quad \text{and} \quad d\overline{B}_t^1 d\overline{B}_t^1 = dt = d\overline{B}_t^2 d\overline{B}_t^2$$

so that (6.76) can be rewritten as follows

$$(6.76) \iff \begin{cases} d\Theta_t^1 &:= -\dfrac{\sin(\Theta_t^1)}{2r(R+r\cos(\Theta_t^1))}\,dt+\dfrac{1}{r}\,dB_t^1 \\ d\Theta_t^2 &= \dfrac{1}{(R+r\cos(\Theta_t^1))}\,dB_t^2 \end{cases}$$

6.5.3 Diffusions on the simplex

We return to the Brownian motion on the orbifold $S/\mathcal{H} = \mathbb{S}^p \cap \mathbb{R}_+^{r=p+1}$ discussed in section 6.2.7. The positive orthant of the sphere is in bijection with the p-simplex

$$\text{Simplex}(p) = \{\theta = (\theta_i)_{1\leq i\leq r} \in \mathbb{R}_+^r \ : \ \sum_{1\leq i\leq r} \theta_i = 1\}$$

One diffeomorphism is given by the square mapping

$$\begin{aligned} x = (x_i)_{1\leq i\leq r}^T \in \mathbb{S}^p \cap \mathbb{R}_+^{r=p+1} &\xrightarrow{\ \Xi\ } \Xi(x) = \left(x_1^2,\ldots,x_r^2\right)^T \in \text{Simplex}(p) \\ \left(\sqrt{\theta_1},\ldots,\sqrt{\theta_r}\right)^T = \Xi^{-1}(\theta) \in \mathbb{S}^p \cap \mathbb{R}_+^{r=p+1} &\xleftarrow{\ \Xi^{-1}\ } \theta = (\theta_i)_{1\leq i\leq r}^T \in \text{Simplex}(p) \end{aligned}$$

Notice that

$$\partial_{x_k}\Xi^i = 2\ 1_{k=i}\ x^i \quad \Longrightarrow \quad \frac{1}{2}\,\partial_{x_k,x_l}\Xi^i = 1_{k,l=i} \tag{6.81}$$

Using (6.22), the projection $\pi(x)$ onto $T_x(\mathbb{S}^p \cap \mathbb{R}_+^{r=p+1})$ is given by

$$\begin{aligned} x = \Xi^{-1}(\theta) \Rightarrow \pi(x) &= Id - \frac{(\partial\varphi)(x)}{\|(\partial\varphi)(x)\|} \frac{(\partial\varphi)(x)^T}{\|(\partial\varphi)(x)\|} \\ &= \begin{pmatrix} 1 - \frac{x_1^2}{\|x\|^2} & -\frac{x_1 x_2}{\|x\|} & -\frac{x_1 x_3}{\|x\|} & \dots & -\frac{x_1 x_r}{\|x\|} \\ -\frac{x_2 x_1}{\|x\|} & 1 - \frac{x_2^2}{\|x\|^2} & -\frac{x_2 x_3}{\|x\|} & \dots & -\frac{x_2 x_r}{\|x\|} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{x_r x_1}{\|x\|} & -\frac{x_r x_2}{\|x\|} & \dots & \dots & 1 - \frac{x_r^2}{\|x\|} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{\theta_1}{\sum_{1\leq j\leq r}\theta_j} & -\frac{\sqrt{\theta_1}\sqrt{\theta_2}}{\sum_{1\leq j\leq r}\theta_j} & -\frac{\sqrt{\theta_1}\sqrt{\theta_3}}{\sum_{1\leq j\leq r}\theta_j} & \dots & -\frac{\sqrt{\theta_1}\sqrt{\theta_r}}{\sum_{1\leq j\leq r}\theta_j} \\ -\frac{\sqrt{\theta_2}\sqrt{\theta_1}}{\sum_{1\leq j\leq r}\theta_j} & 1 - \frac{\theta_2}{\sum_{1\leq j\leq r}\theta_j} & -\frac{\sqrt{\theta_2}\sqrt{\theta_3}}{\sum_{1\leq j\leq r}\theta_j} & \dots & -\frac{\sqrt{\theta_2}\sqrt{\theta_r}}{\sum_{1\leq j\leq r}\theta_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{\sqrt{\theta_r}\sqrt{\theta_1}}{\sum_{1\leq j\leq r}\theta_j} & -\frac{\sqrt{\theta_r}\sqrt{\theta_2}}{\sum_{1\leq j\leq r}\theta_j} & \dots & \dots & 1 - \frac{\theta_r}{\sum_{1\leq j\leq r}\theta_j} \end{pmatrix} \end{aligned}$$

In addition, using (6.23) the curvature vector $\mathbb{H}$ on the sphere is given by

$$x = \Xi^{-1}(\theta) \Rightarrow \mathbb{H}^i(x) = p\frac{x_i}{x^T x} = p\ \frac{\sqrt{\theta_i}}{\sum_{1\leq l\leq r}\theta_l}$$

We let X_t be the Brownian motion on the positive orthant defined in (6.35). We recall that

$$dX_t^k = -\frac{1}{2}\ \mathbb{H}^k(X_t)\ dt + \sum_{1\leq j\leq r} \pi_j^k(X_t)\ dB_t^j \Rightarrow dX_t^k dX_t^k = \left[\sum_{1\leq j\leq r} \pi_j^k \pi_j^l\right](X_t)\ dt$$

Applying Ito formula to

$$\Theta_t = \Xi(X_t) \quad \Leftrightarrow \quad \Xi^{-1}(\Theta_t) = X_t = \left(\sqrt{\Theta_t^1}, \dots, \sqrt{\Theta_t^r}\right)$$

we find that

$$\begin{aligned} d\Theta_t^i &= \sum_{1\leq k\leq r} \left(\partial_{x_k}\Xi^i\right)\left(\Xi^{-1}(\Theta_t)\right) \left[-\frac{1}{2}\ \mathbb{H}^k(\Xi^{-1}(\Theta_t))\ dt + \sum_{1\leq j\leq r} \pi_j^k(\Xi^{-1}(\Theta_t))\ dB_t^j\right] \\ &\quad + \frac{1}{2}\sum_{1\leq k,l\leq r} \left(\partial_{x_k,x_l}\Xi^i\right)\left(\Xi^{-1}(\Theta_t)\right) \left[\sum_{1\leq j\leq r} \pi_j^k \pi_j^l\right](\Xi^{-1}(\Theta_t))\ dt \\ &= 2\sqrt{\Theta_t^i}\left[-\frac{1}{2}\ \mathbb{H}^i(\Xi^{-1}(\Theta_t))\ dt + \sum_{1\leq j\leq r} \pi_j^i(\Xi^{-1}(\Theta_t))\ dB_t^j\right] + \left[\sum_{1\leq k\leq r} \pi_k^i \pi_k^i\right](\Xi^{-1}(\Theta_t))\ dt \end{aligned}$$

The last assertion is a direct consequence of (6.81). To take the final step, we observe that

$$\begin{aligned} \sum_{1\leq k\leq r} \left(\pi_k^i(\Xi^{-1}(\theta))\right)^2 &= \left(1 - \frac{\theta_i}{\sum_{1\leq j\leq r}\theta_j}\right)^2 + \frac{\theta_i}{\sum_{1\leq j\leq r}\theta_j} \sum_{1\leq k\leq r,\ k\neq i} \frac{\theta_k}{\sum_{1\leq j\leq r}\theta_j} \\ &= 1 - 2\ \frac{\theta_i}{\sum_{1\leq j\leq r}\theta_j} + \frac{\theta_i}{\sum_{1\leq j\leq r}\theta_j} \sum_{1\leq k\leq r} \frac{\theta_k}{\sum_{1\leq j\leq r}\theta_j} \\ &= 1 - \frac{\theta_i}{\sum_{1\leq j\leq r}\theta_j} \end{aligned}$$

and

$$\begin{aligned}\sum_{1\leq j\leq r}\pi_j^i(\Xi^{-1}(\Theta_t))\ dB_t^j &= \left(1-\frac{\Theta_t^i}{\sum_{1\leq j\leq r}\Theta_t^j}\right)\ dB_t^i-\sum_{1\leq j\leq r,\ j\neq i}\frac{\sqrt{\Theta_t^i}\sqrt{\Theta_t^j}}{\sum_{1\leq j\leq r}\Theta_t^j}\ dB_t^j\\ &= dB_t^i-\sum_{1\leq j\leq r}\frac{\sqrt{\Theta_t^i}\sqrt{\Theta_t^j}}{\sum_{1\leq j\leq r}\Theta_t^j}\ dB_t^j\end{aligned}$$

We conclude that

$$\begin{aligned}d\Theta_t^i &= \left(-p\ \frac{\Theta_t^i}{\sum_{1\leq l\leq r}\Theta_t^l}+\left(1-\frac{\Theta_t^i}{\sum_{1\leq j\leq r}\Theta_t^j}\right)\right)\ dt\\ &\qquad +2\sqrt{\Theta_t^i}\ \left[dB_t^i-\sum_{1\leq j\leq r}\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq j\leq r}\Theta_t^j}\ dB_t^j\right]\\ &= \left(1-r\ \frac{\Theta_t^i}{\sum_{1\leq j\leq r}\Theta_t^j}\right)\ dt+2\sqrt{\Theta_t^i}\ \left[dB_t^i-\sum_{1\leq j\leq r}\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq j\leq r}\Theta_t^j}\ dB_t^j\right]\end{aligned}$$

It is instructive to notice that

$$\begin{aligned}&d\Theta_t^i d\Theta_t^j\\ &= 4\sqrt{\Theta_t^i\Theta_t^i}\ \left[\left(1_{i=j}-\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq l\leq r}\Theta_t^l}\right)+\left(-\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq j\leq r}\Theta_t^j}+\sum_{1\leq k\leq r}\frac{\Theta_t^k}{\sum_{1\leq l\leq r}\Theta_t^l}\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq l\leq r}\Theta_t^l}\right)\right]\ dt\\ &= 4\sqrt{\Theta_t^i\Theta_t^i}\ \left[\left(1_{i=j}-\frac{\sqrt{\Theta_t^i}\ \sqrt{\Theta_t^j}}{\sum_{1\leq l\leq r}\Theta_t^l}\right)\right]\ dt=4\Theta_t^i\ \left(1_{i=j}-\frac{\Theta_t^j}{\sum_{1\leq l\leq r}\Theta_t^l}\right)dt\end{aligned}$$

6.6 Some analytical aspects

6.6.1 Geodesics and the exponential map

The distance between two states $x, y \in S$ is defined in a chart coordinate by the formula

$$d(x,y)=\inf\ \int_a^b \left\|\dot{c}(t)\right\|_{g(c(t))}\ dt$$

where the infimum is taken over all parametric curves $c : t \in [a,b] \mapsto c(t) \in S_\psi$ s.t. $\phi(c(a)) = x$, and $\phi(c(b)) = y$, and

$$\begin{aligned}\left\|\dot{c}(t)\right\|_{g(c(t))}^2 &:= \left\langle\dot{c}(t),\dot{c}(t)\right\rangle_{g(c(t))}\\ &= \sum_{1\leq i,j\leq p} g_{i,j}(c(t))\ \dot{c}^i(t)\,\dot{c}^j(t)\\ &= \left\langle\sum_{1\leq i\leq p}\dot{c}^i(t)\ (\partial_{\theta_i}\psi)_{c(t)},\sum_{1\leq j\leq p}\dot{c}^j(t)\ (\partial_{\theta_i}\psi)_{c(t)}\right\rangle\\ &= \left\langle\dot{C}(t),\dot{C}(t)\right\rangle=\left\|\dot{C}(t)\right\|^2\quad\text{with}\quad C(t)=\psi(c(t))\end{aligned}$$

To better understand this definition, we simply notice that

$$\|c(t+dt)-c(t)\|_{g(c(t))} \simeq \left\|\dot{c}\,(t)\right\|_g \; dt \Rightarrow \sum_{a\leq t\leq b} \|c(t+dt)-c(t)\|_{g(c(t))} \; dt \simeq \int_a^b \left\|\dot{c}\,(t)\right\|_{g(c(t))} \; dt$$

The energy of a given curve c is given by

$$\mathcal{E}(c) \;=\; \frac{1}{2}\int_a^b \left\|\dot{c}\,(t)\right\|^2_{g(c(t))} \; dt = \int_a^b L\left(c(t),\dot{c}\,(t)\right) \; dt$$

with the Lagrangian

$$L(c,\dot{c}) := \frac{1}{2} \; \left\|\dot{c}\right\|^2_{g(c)} = \sum_{1\leq i,j\leq p} g_{i,j}(c) \;\; \dot{c}^i \dot{c}^j$$

To find the extremal curves, we let $c_\epsilon(t) = c(t) + \epsilon c'(t)$ and ϵ-perturbation of c, with some curve $c'(t)$ s.t. $c'(a) = 0 = c'(b)$. For any $t \in [a,b]$, we have

$$\frac{d}{d\epsilon} L(c_\epsilon(t), \dot{c}_\epsilon\,(t))_{|\epsilon=0} = \left\langle (\partial_c L)(c(t),\dot{c}\,(t)), c'(t)\right\rangle + \left\langle (\partial_{\dot{c}} L)(c(t),\dot{c}\,(t)), \dot{c}'\,(t)\right\rangle$$

with the gradients

$$(\partial_c L) = \begin{bmatrix} (\partial_{c^1} L) \\ \vdots \\ (\partial_{c^p} L) \end{bmatrix} \quad \text{and} \quad (\partial_{\dot{c}} L) = \begin{bmatrix} (\partial_{\dot{c}^1} L) \\ \vdots \\ (\partial_{\dot{c}^p} L) \end{bmatrix}$$

An integration by part w.r.t. the time parameter yields

$$\int_a^b \left\langle (\partial_{\dot{c}} L)(c(t),\dot{c}\,(t)), \dot{c}'\,(t)\right\rangle dt = \left[\left\langle (\partial_{\dot{c}} L)(c(t),\dot{c}\,(t)), c'(t)\right\rangle\right]_a^b - \int_a^b \left\langle \frac{d}{dt}\left[(\partial_{\dot{c}} L)(c(t),\dot{c}\,(t))\right], c'(t)\right\rangle dt$$

This implies that for any perturbation c' we have

$$\frac{d}{d\epsilon}\mathcal{E}(c_\epsilon)_{|\epsilon=0} = \int_a^b \left\langle (\partial_c L)(c(t),\dot{c}\,(t)) - \frac{d}{dt}\left[(\partial_{\dot{c}} L)(c(t),\dot{c}\,(t))\right], c'(t)\right\rangle dt$$

from which we conclude that the extremal curves satisfy the differential Lagrange equation

$$(\partial_c L)(c(t),\dot{c}\,(t)) = \frac{d}{dt}\left[(\partial_{\dot{c}} L)(c(t),\dot{c}\,(t))\right], \tag{6.82}$$

In our context, we have

$$\partial_{c^k} L = \sum_{1\leq i,j\leq p} \left(\partial_{\theta_k} g_{i,j}\right)(c) \;\; \dot{c}^i \;\; \dot{c}^j$$

and

$$\partial_{\dot{c}^k} L = 2 \; \textstyle\sum_{1\leq i\leq p} \; g_{k,i} \;\; \dot{c}^i$$

$$\Rightarrow \tfrac{d}{dt}\left[(\partial_{\dot{c}^k} L)(c(t),\dot{c}\,(t))\right] = 2 \; \textstyle\sum_{1\leq i,j\leq p} \; \left(\partial_{\theta_j} g_{k,i}\right)(c(t)) \;\; \dot{c}^i\,(t)\,\dot{c}^j\,(t) + 2 \; \textstyle\sum_{1\leq i\leq p} \; g_{k,i}(c(t)) \;\; \ddot{c}^i\,(t)$$

Thus, the Lagrange equations take the following form

$$(6.82) \Leftrightarrow \forall 1 \leq k \leq p \qquad \sum_{1\leq i\leq p} g_{k,i}(c(t)) \;\; \ddot{c}^i\,(t)$$

$$= \sum_{1\leq i,j\leq p} \left[\frac{1}{2}\left(\partial_{\theta_k} g_{i,j}\right)(c(t)) - \left(\partial_{\theta_j} g_{k,i}\right)(c(t))\right] \;\; \dot{c}^i\,(t) \;\; \dot{c}^j\,(t)$$

We conclude that

$$\begin{aligned}\ddot{c}^m(t) &= \sum_{1\le i,k\le p} g^{m,k}(c(t))\; g_{k,i}(c(t))\;\; \ddot{c}^i(t)\\ &= \sum_{1\le i,j\le p}\left(\sum_{1\le k\le p} g^{m,k}\left[\frac{1}{2}\,\partial_{\theta_k} g_{i,j} - \partial_{\theta_j} g_{k,i}\right]\right)(c(t))\;\; \dot{c}^i(t)\;\; \dot{c}^j(t)\end{aligned}$$

Next, we express this formula in terms of the Christoffel symbol $\Gamma^k_{i,j}$ introduced in (6.58). Firstly, we notice that

$$\begin{aligned}&\left\langle \delta_{\theta_l}\psi, \delta_{\theta_i,\theta_j}\psi\right\rangle = \delta_{\theta_i}\left\langle \delta_{\theta_l}\psi, \delta_{\theta_j}\psi\right\rangle - \left\langle \delta_{\theta_i\theta_l}\psi, \delta_{\theta_j}\psi\right\rangle = \delta_{\theta_i} g_{l,j} - \left\langle \delta_{\theta_i\theta_l}\psi, \delta_{\theta_j}\psi\right\rangle\\ \Rightarrow\quad \Gamma^k_{i,j} &= \Gamma^k_{j,i}\\ &= \textstyle\sum_{1\le l\le p}\; g^{k,l}\;\left\langle \delta_{\theta_l}\psi, \delta_{\theta_i,\theta_j}\psi\right\rangle\\ &= \textstyle\sum_{1\le l\le p}\; g^{k,l}\;\delta_{\theta_i} g_{l,j} - \sum_{1\le l\le p}\; g^{k,l}\;\left\langle \delta_{\theta_l\theta_i}\psi, \delta_{\theta_j}\psi\right\rangle\end{aligned}$$

Thus, for any symmetric functionals $f^{i,j} = f^{j,i}$, $1 \le i, j \le p$, on S_ψ we have

$$\begin{aligned}\sum_{1\le i,j\le p}\Gamma^m_{i,j}\; f^{i,j} &= \sum_{1\le i,j\le p} f^{i,j}\sum_{1\le k\le p} g^{m,k}\;\delta_{\theta_i} g_{k,j} - \sum_{1\le i,j\le p} f^{i,j}\sum_{1\le k\le p} g^{m,k}\;\left\langle \delta_{\theta_k\theta_i}\psi, \delta_{\theta_j}\psi\right\rangle\\ &= -\sum_{1\le i,j\le p}\;\sum_{1\le k\le p} g^{m,k}\;\left[\frac{1}{2}\,\partial_{\theta_k} g_{i,j} - \;\delta_{\theta_i} g_{k,j}\right]\; f^{i,j}\end{aligned}$$

We conclude that

$$\forall 1 \le m \le p \qquad \ddot{c}^m(t) = -\sum_{1\le i,j\le p}\;\Gamma^m_{i,j}\;\; \dot{c}^i(t)\;\; \dot{c}^j(t)$$

The solution of these equations gives a curve that minimizes the distances between two states $\phi(x)$ and $\phi(y)$ in the parameter space. These curves $c(t)$ and their mapping $C(t) = \psi(c(t))$ into the manifold S are called the geodesics.

It is instructive to observe that the velocity vector $C(t) = \psi(c(t))$ of a given curve on S is given by the formula

$$\frac{dC}{dt}(t) = \sum_{1\le i\le p}\left(\partial_{\theta_i}\psi\right)(c(t))\;\; \dot{c}^i_t$$

Thus, its acceleration takes the form

$$\frac{d^2C}{dt^2}(t) = \sum_{1\le i\le p}\left(\partial_{\theta_i}\psi\right)(c(t))\;\; \ddot{c}^i_t + \sum_{1\le i,j\le p}\left(\partial_{\theta_j,\theta_i}\psi\right)(c(t))\;\; \dot{c}^i_t\;\;\; \dot{c}^j_t$$

Taking the orthogonal projection on the tangent plane $T_{C(t)}(S)$ we have

$$\begin{aligned}\pi(C(t))\left(\frac{d^2C}{dt^2}(t)\right) &= \sum_{1\leq i\leq p}\left(\partial_{\theta_i}\psi\right)(c(t))\ \ddot{c}_t^i + \sum_{1\leq i,j\leq p}\pi\left[\left(\partial_{\theta_j,\theta_i}\psi\right)(c(t))\right]\ \dot{c}_t^i\ \dot{c}_t^j \\ &= \sum_{1\leq m\leq p}\left[\ \ddot{c}_t^m + \sum_{1\leq i,j\leq p}\Gamma_{i,j}^m(c(t))\ \dot{c}_t^i\ \dot{c}_t^j\right]\ \left(\partial_{\theta_m}\psi\right)(c(t)) = 0\end{aligned}$$

from which we find that

$$\forall 1\leq m\leq p \qquad \ddot{c}_t^m + \sum_{1\leq i,j\leq p}\Gamma_{i,j}^m(c(t))\ \dot{c}_t^i\ \dot{c}_t^j = 0$$

This shows that the acceleration vector of the geodesics is orthogonal to the tangent place $T(S)$. In other words, the speed geodesic vector $\dot{c}\,(t)$ is parallel to the curve $c(t)$ (cf. (6.61))

By the existence and uniqueness theorem of solutions of ordinary differential equations, given a tangent vector field $W(x)\in T_x(S)$, for any $x=\psi(\theta)\in S$ there exists an unique geodesics $C_x(t)$ with velocity vector $W(x)=\frac{dC_x}{dt}(0)$ at the origin. The geodesics $c_{\phi(x)}(t)=\phi(C_x(t))$ and $C_x(t)$ associated with a given velocity vector $V_\phi(x)$ and $W_x=(d\psi)_{\phi(x)}(V_{\phi(x)})$ are denoted in terms of exponential maps

$$C_x(t) := Exp_x(tW) \quad \text{and} \quad c_\theta(t) := Exp_\theta(tV)$$

6.6.2 A Taylor expansion

Given a smooth function f on the parameter space S_ψ, we have

$$\begin{aligned}\frac{d}{dt}f\left(c_\theta(t)\right) &= \sum_{1\leq i\leq p}\left(\partial_{\theta_i}f\right)\left(c_\theta(t)\right)\ \frac{dc_\theta^i}{dt}(t) \\ &= \sum_{1\leq i\leq p}\left(\partial_{\theta_i}f\right)\left(c_\theta(t)\right)\ V^i\left(c_\theta(t)\right) = \partial_V(f)\left(c_\theta(t)\right) \\ &= \left\langle V\left(c_\theta(t)\right),\left(\partial f\right)\left(c_\theta(t)\right)\right\rangle = \left\langle V\left(c_\theta(t)\right),\left(\nabla_g f\right)\left(c_\theta(t)\right)\right\rangle_{g(c_\theta(t))}\end{aligned}$$

Thus, for $t=0$ we have

$$\frac{d}{dt}f\left(Exp_\theta(tV)\right)_{|t=0} = \left\langle V\left(\theta\right),\left(\nabla_g f\right)\left(\theta\right)\right\rangle_{g(\theta)}$$

In much the same way, we have

$$\begin{aligned}\frac{d^2}{dt^2}f\left(c_\theta(t)\right) &= \sum_{1\leq i,j\leq p}\left(\partial_{\theta_i,\theta_j}f\right)\left(c_\theta(t)\right)\ V^i\left(c_\theta(t)\right)V^j\left(c_\theta(t)\right) + \sum_{1\leq i\leq p}\left(\partial_{\theta_i}f\right)\left(c_\theta(t)\right)\ \frac{d^2c_\theta^i}{dt^2}(t) \\ &= \sum_{1\leq k,l\leq p}\left[\left(\partial_{\theta_k,\theta_l}f\right)\left(c_\theta(t)\right) - \sum_{1\leq i\leq p}\Gamma_{k,l}^i(c(t))\ \left(\partial_{\theta_i}f\right)\left(c_\theta(t)\right)\right]\ V^k\left(c_\theta(t)\right)V^l\left(c_\theta(t)\right)\end{aligned}$$

Thus, for $t=0$ we have

$$\frac{d^2}{dt^2}f\left(Exp_\theta(tV)\right)_{|t=0} = V^T(\theta)\left(\nabla_g^2 f\right)(\theta)\ V^T(\theta) = \left\langle V(\theta),\left(\nabla_g^2 f\right)(\theta)\ V(\theta)\right\rangle_{g(\theta)}$$

For regular vector fields V this yields the Taylor expansion

$$f\left(Exp_\theta(tV)\right) = f(\theta) + t\ \left\langle V\left(\theta\right), \left(\nabla_g f\right)\left(\theta\right)\right\rangle_{g(\theta)} + \frac{t^2}{2}\ \left\langle V(\theta), \left(\nabla_g^2 f\right)\left(\theta\right)\ V(\theta)\right\rangle_{g(\theta)} + \mathrm{O}(t^3)$$

or equivalently

$$f\left(Exp_\theta(V)\right) = f(\theta) + \left\langle V\left(\theta\right), \left(\nabla_g f\right)\left(\theta\right)\right\rangle_{g(\theta)} + \frac{1}{2}\ \left\langle V(\theta), \left(\nabla_g^2 f\right)\left(\theta\right)\ V(\theta)\right\rangle_{g(\theta)} + \mathrm{O}\left(\|V\|^3\right)$$

Letting $F = f \circ \phi$, and using the fact that $Exp_\theta(V) = \phi\left(Exp_x(W)\right)$ when $\theta = \phi(x)$ and $W = (d\psi)_\phi(V_\phi)$, the above formula takes the form

$$F\left(Exp_x(W)\right) = F(x) + \left\langle W\left(x\right), \nabla F\left(x\right)\right\rangle + \frac{1}{2}\ \left\langle W(x), (\nabla^2 F)(x) W(x)\right\rangle + \mathrm{O}\left(\|W\|^3\right)$$

6.6.3 Integration on manifolds

The volume measure on the manifold

Heuristically, for manifolds S in dimension

$$p = 1 = \dim(T_x(S)) = \dim(\mathrm{Vect}((\partial_{\theta_i}\psi)_\phi(\theta))$$

the volume element $\mu_S(dx)$ at some state $x = \psi(\theta)$ reduces the length $\mathrm{length}_S(\psi(\delta\theta))$ of the ψ-image curve $\psi(\delta\theta)$ of an infinitesimal interval

$$\delta\theta := [\theta, \theta + d\theta] \in S_\psi \subset \mathbb{R}$$

That is, we have that

$$\psi(\delta\theta) \simeq \psi(\theta + d\theta) - \psi(\theta) \simeq \left(\partial_\theta \psi\right)\left(\theta\right)\ d\theta$$

so that

$$\begin{aligned} \mu_S(dx) &= \mathrm{length}_S(\psi(\delta\theta)) \simeq \|\psi(\theta + d\theta) - \psi(\theta)\| \\ &\simeq \left\|\left(\partial_\theta \psi\right)\left(\theta\right)\right\|\ d\theta = \sqrt{\left\langle \left(\partial_\theta \psi\right)\left(\theta\right), \left(\partial_\theta \psi\right)\left(\theta\right)\right\rangle}\ d\theta \end{aligned}$$

More rigorously, for any function F with compact support we have

$$\int_S F(x)\ \mu_S(dx) = \int_{S_\psi} f(\theta)\ \sqrt{\left\langle \left(\partial_\theta \psi\right)\left(\theta\right), \left(\partial_\theta \psi\right)\left(\theta\right)\right\rangle}\ d\theta \quad \text{with} \quad f = F \circ \psi$$

In larger dimensions, the ψ-image $\psi(\delta\theta)$ of a cell $\delta\theta = \prod_{1\leq i\leq p}[\theta_i, \theta_i + d\theta_i] \in S_\psi \subset \mathbb{R}^p$ is given by

$$\begin{aligned} \psi(\delta\theta) &\simeq \psi(\theta + d\theta) - \psi(\theta) \simeq \sum_{1\leq i\leq p} \left(\partial_{\theta_i}\psi\right)\left(\theta\right)\ d\theta_i \\ &= (d\psi)\left(\left\{\sum_{1\leq i\leq p} \epsilon_i\ e_i\ :\ \epsilon_i \in [0, d\theta_i]\right\}\right) = (d\psi)\left(\prod_{1\leq i\leq p}[0, d\theta_i]\right) \end{aligned}$$

with the unit vectors e_i in $\mathbb{R}^p$, $1 \leq i \leq p$. We recall that $(d\psi)(e_i) = \partial_{\theta_i}$, for any $1 \leq i \leq p$. On the other hand, by the change of variables formula

$$\text{Vol}\left[(d\psi)\left(\prod_{1\leq i\leq p}[0,d\theta_i]\right)\right] \simeq \det\left((d\psi)(\theta)\right) \underbrace{\prod_{1\leq i\leq p} d\theta_i}_{=d\theta} = \det\left((d\psi)(\theta)\right)\ d\theta$$

Recalling that $\sqrt{\det(A^T A)} = \det(A)$, and $A_{i,j} = \langle Ae_i, e_j\rangle$ for any $(p\times p)$-matrix A, we have

$$\begin{aligned}
\det\left((d\psi)(\theta)\right) &= \sqrt{\det\left((d\psi)(\theta)^T(d\psi)(\theta)\right)} \\
&= \sqrt{\det\left(\langle(d\psi)(\theta)^T(d\psi)(\theta)e_i, e_j\rangle\right)_{1\leq i,j\leq p}} \\
&= \sqrt{\det\left(\langle(d\psi)(\theta)e_i, (d\psi)(\theta)e_j\rangle\right)_{1\leq i,j\leq p}} \\
&= \sqrt{\det\left(\left\langle\left(\partial_{\theta_i}\psi\right)(\theta), \left(\partial_{\theta_j}\psi\right)(\theta)\right\rangle\right)_{1\leq i,j\leq p}} = \sqrt{\det\left(g(\theta)\right)}
\end{aligned}$$

In summary, for any function F with compact support we have

$$\int_S F\ d\mu_S := \int_S F(x)\ \mu_S(dx) = \int_{S_\psi} f(\theta)\ \sqrt{\det\left(g(\theta)\right)}\ d\theta \quad \text{with} \quad f = F\circ\psi$$

If we set

$$\mu_g(d\theta) = \sqrt{\det\left(g(\theta)\right)}\ d\theta$$

the above formulae can be rewritten in a more synthetic form

$$\mu_S(F) := \int_S F\ d\mu_S = \int_{S_\psi} f(\theta)\ \mu_g(d\theta) := \mu_g\left(f\right)$$

The divergence theorem

We consider the push forward $W = (d\psi)_\phi(V_\phi)$ and $F = f\circ\psi$ of a smooth vector field V and a smooth function f on S with compact support. By construction (cf. (6.45)), we have

$$W \in T(S) \Rightarrow \langle W, \nabla F\rangle = \langle W, \partial F\rangle = \partial_W F = (\partial_V f)\circ\phi = \langle V, \partial f\rangle\circ\phi$$

and

$$\begin{aligned}
\int_S \langle W, \nabla F\rangle\ d\mu_S &= \int_{S_\psi} \langle V(\theta), (\partial f)(\theta)\rangle\ \sqrt{\det\left(g(\theta)\right)}\ d\theta \\
&= \int_{S_\psi} \langle V(\theta), (\nabla_g f)(\theta)\rangle_{g(\theta)}\ \mu_g(d\theta)
\end{aligned}$$

with

$$\mu_g(d\theta) = \sqrt{\det\left(g(\theta)\right)}\ d\theta$$

Using the fact that

$$\langle V(\theta), (\partial f)(\theta)\rangle = \sum_{1\leq i\leq p} V^i(\theta)\ \partial_{\theta_i}(f)(\theta)$$

by a simple integration by part formula, we prove that

$$\begin{aligned}
&\int_{S_\psi} \langle V(\theta), (\partial f)(\theta)\rangle \ \sqrt{\det\left(g(\theta)\right)}\ d\theta \\
&= \sum_{1\le i\le p} \int_{S_\psi} V^i(\theta)\ \partial_{\theta_i}(f)(\theta)\ \sqrt{\det\left(g(\theta)\right)}\ d\theta \\
&= -\sum_{1\le i\le p} \int_{S_\psi} f(\theta)\ \partial_{\theta_i}\left[\sqrt{\det\left(g(\theta)\right)}\ V^i(\theta)\ \right]\ d\theta \\
&= -\int_{S_\psi} f(\theta)\ \underbrace{\frac{1}{\sqrt{\det\left(g(\theta)\right)}} \sum_{1\le i\le p} \partial_{\theta_i}\left[\sqrt{\det\left(g(\theta)\right)}\ V^i(\theta)\ \right]}_{=\mathrm{div}_g(V)(\theta)}\ \sqrt{\det\left(g(\theta)\right)}\ d\theta
\end{aligned}$$

This implies that

$$\int_S \langle W, \nabla F\rangle\ d\mu_S = -\int_{S_\psi} f(\theta)\ \mathrm{div}_g(V)(\theta)\ \sqrt{\det\left(g(\theta)\right)}\ d\theta = -\int_{S_\psi} f(\theta)\ \mathrm{div}_g(V)(\theta)\ \mu_g(d\theta)$$

On the other hand, using (6.54) we have

$$\mathrm{div}_g(V)(\theta) = \mathrm{tr}\left(\nabla W\right)_\psi = \mathrm{div}\left(W\right)\circ\psi$$

from which we conclude that

$$\int_S F\ \mathrm{div}\left(W\right)\ d\mu_S = -\int_S \langle W, \nabla F\rangle\ d\mu_S \tag{6.83}$$

We quote a series of direct consequence of this integration by part formula:

- Choosing $F = 1$ we have
$$\int_S \mathrm{div}\left(W\right)\ d\mu_S = 0$$
- Choosing $W = F_1\ \nabla F_2$ and $F = F_3$, we have
$$\begin{aligned}
\int_S \mathrm{div}\left(F_1\ \nabla F_2\right)\ F_3\ d\mu_S &= -\int_S \langle F_1\ \nabla F_2, \partial F_3\rangle\ d\mu_S \\
&= -\int_S F_1\ \langle \nabla F_2, \partial F_3\rangle\ d\mu_S = -\int_S F_1\ \langle \nabla F_2, \nabla F_3\rangle\ d\mu_S
\end{aligned}$$
The r.h.s. assertion comes from the fact that
$$\nabla F_2 \in T(S) \implies \langle \nabla F_2, \partial F_3\rangle = \langle \nabla F_2, \pi\left(\partial F_3\right)\rangle = \langle \nabla F_2, \nabla F_3\rangle$$

- Choosing $F_1 = 1$ in the above formula, and combining (6.50) with (6.55), we prove the first type Green formula; that is, for any smooth functions (F_1, F_2) with (at least one with) compact support

$$\int_S \Delta(F_1)\ F_2\ d\mu_S = \int_S \mathrm{div}\,(\nabla F_1)\ F_2\ d\mu_S = -\int_S \langle \nabla F_1, \nabla F_2 \rangle\ d\mu_S = \int_S F_1\ \Delta(F_2)\ d\mu_S$$

$$=$$

$$\int_{S_\psi} \Delta_g(f_1)\ f_2\ d\mu_g = \int_S \mathrm{div}_g\,(\nabla_g f_1)\ f_2\ d\mu_g = -\int_{S_\psi} \langle \nabla_g f_1, \nabla_g f_2 \rangle_g\ d\mu_g = \int_{S_\psi} f_1\ \Delta_g(f_2)\ d\mu_g$$

- Choosing $F_2 = 1$ in the above formula, we find that for any smooth function F with compact support

$$\int_S \Delta(F)\ d\mu_S = 0$$

6.7 Some prototype manifolds

6.7.1 The Circle

The prototype of hypermanifold is the unit circle $S = \mathbb{S}^1 \subset \mathbb{R}^2 \ni x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ described as the null level set $S = \varphi^{-1}(0)$ of the function

$$\varphi(x) = (x_1^2 + x_2^2 - 1)/2 \Rightarrow (\partial\varphi)(x) = x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

The orthogonal projection $\pi_\perp$ onto the normal axis $T_x^\perp(S) = \mathrm{Vect}\,((\partial\varphi)(x))$ at $x \in S$ is given by the formula

$$\pi_\perp(x) = \frac{(\partial\varphi)(x)\,(\partial\varphi)^T(x)}{(\partial\varphi)(x)^T\,(\partial\varphi)(x)} = \frac{xx^T}{x^T x} = \frac{1}{x_1^2 + x_2^2}\begin{pmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{pmatrix}$$

and the orthogonal projection on $T_x(S)$ is defined by $\pi(x) = Id - \pi_\perp(x)$. The (mean) curvature vector $\mathbb{H}$ defined by (6.21) on the circle is simply given by

$$\forall x \neq 0 \qquad \mathbb{H}(x) = \left[\sum_{1\leq m\leq 2} \partial_{x_m}\left(\frac{x_m}{\sqrt{x_1^2 + x_2^2}}\right)\right]\ \frac{x}{\sqrt{x_1^2 + x_2^2}} = \frac{x}{\|x\|^2}$$

We check this claim using the fact that

$$\partial_{x_1}\left(\frac{x_1}{\sqrt{x_1^2 + x_2^2}}\right) = \frac{1}{\sqrt{x_1^2 + x_2^2}}\left[1 - \frac{x_1^2}{(x_1^2 + x_2^2)}\right]$$
$$\Rightarrow \mathrm{div}_\perp\left(\frac{\partial\varphi}{\|\partial\varphi\|^2}\right) = \sum_{1\leq m\leq 2} \partial_{x_m}\left(\frac{x_m}{\sqrt{x_1^2 + x_2^2}}\right) = \frac{1}{\sqrt{x_1^2 + x_2^2}}$$

The circle $S - \{(1,0)\}$ can be parametrized by the polar angle mapping $\psi\ :\ \theta \in]0, 2\pi[\mapsto S - \{(1,0)\}$

$$\psi(\theta) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \quad \Longrightarrow \quad (\partial_\theta\psi)(\theta) = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}$$

so that

$$T_x(S) = \mathrm{Vect}\left((\partial_\theta\psi)_{\phi(x)}\right) \quad \text{with} \quad (\partial_\theta\psi)_{\phi(x)} = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}_{\theta=\phi(x)} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$$

The Riemannian metric on $S_\psi = \mathbb{R} \Rightarrow T(S_\psi) = \mathbb{R} = \text{Vect}(1)$ reduces to

$$g(\theta) = \langle (\partial_\theta \psi)(\theta), (\partial_\theta \psi)(\theta) \rangle = 1 = g(\theta)^{-1} \Rightarrow (\nabla \phi)_\psi = \partial_\theta \psi \tag{6.84}$$

Using (6.19) and (6.13), for any smooth function F on $S \ni x$ we have

$$\text{div}_\perp (\partial \varphi) \quad = \quad \partial_{x_1} (\partial_{x_1} \varphi) + \partial_{x_2} (\partial_{x_2} \varphi) = 2$$

and therefore

$$\frac{1}{2} \Delta F \quad = \quad \text{tr} \left(\pi \partial^2 F \right) - \langle \partial \varphi, \partial F \rangle$$

In addition, we have

$$(\partial_{\theta,\theta} \psi)(\theta) = - \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \in T^\perp(S) \Rightarrow \Gamma^1_{1,1} = 0 \quad \text{and} \quad (\Delta \phi)_\psi = 0 \tag{6.85}$$

The geodesics $c_\theta(t) := Exp_\theta(tV)$, with $V(\theta) \in \mathbb{R}$ are given by

$$\begin{aligned} \ddot{c}_\theta (t) = 0 \quad &\Rightarrow \quad \dot{c}_\theta (t) = V(\theta) \\ &\Rightarrow \quad c_\theta(t) = t\, V(\theta) + \theta \Rightarrow C_x(t) = \psi(c_\theta(t)) = \begin{pmatrix} \cos(t\, V(\theta) + \theta) \\ \sin(t\, V(\theta) + \theta) \end{pmatrix} \end{aligned}$$

6.7.2 The 2-Sphere

The unit sphere $S = \mathbb{S}^2 \subset \mathbb{R}^3 \ni x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ described as the null level set $S = \varphi^{-1}(0)$ of the function

$$\varphi(x) = (x_1^2 + x_2^2 + x_3^2 - 1)/2 \Rightarrow (\partial \varphi)(x) = x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Notice that $(\partial \varphi)(x)$ is the unit normal at any state $x \in S$. Thus, the orthogonal projection $\pi_\perp$ onto the normal axis $T_x^\perp(S) = \text{Vect}\left((\partial \varphi)(x)\right)$ at $x \in S$ is given by the formula

$$\pi_\perp(x) = (\partial \varphi)(x) \, (\partial \varphi)^T (x) = x x^T = \begin{pmatrix} x_1^2 & x_1 x_2 & x_1 x_3 \\ x_2 x_1 & x_2^2 & x_2 x_3 \\ x_3 x_1 & x_2 x_2 & x_3^2 \end{pmatrix}$$

and the orthogonal projection on $T_x(S)$ is defined by

$$\pi(x) = Id - \pi_\perp(x) = \begin{pmatrix} 1 - x_1^2 & -x_1 x_2 & -x_1 x_3 \\ -x_2 x_1 & 1 - x_2^2 & -x_2 x_3 \\ -x_3 x_1 & -x_2 x_2 & 1 - x_3^2 \end{pmatrix}$$

The sphere S can be parametrized by the spherical coordinates mapping $\psi : \theta = (\theta_1, \theta_2) \in]0, \pi[\times]0, 2\pi[\mapsto S$

$$\psi(\theta) = \begin{pmatrix} \sin(\theta_1) \cos(\theta_2) \\ \sin(\theta_1) \sin(\theta_2) \\ \cos(\theta_1) \end{pmatrix} = (\partial \varphi)_\psi(\theta)$$

The first coordinate θ_1 is called the colatitude angle, and the second one θ_2 is called the azimuthal angle. We have

$$\partial_{\theta_1} \psi(\theta) = \begin{pmatrix} \cos(\theta_1) \cos(\theta_2) \\ \cos(\theta_1) \sin(\theta_2) \\ -\sin(\theta_1) \end{pmatrix} \quad \text{and} \quad \partial_{\theta_2} \psi(\theta) = \begin{pmatrix} -\sin(\theta_1) \sin(\theta_2) \\ \sin(\theta_1) \cos(\theta_2) \\ 0 \end{pmatrix} = -\sin(\theta_1) \begin{pmatrix} \sin(\theta_2) \\ -\cos(\theta_2) \\ 0 \end{pmatrix}$$

so that

$$\partial_{\theta_1}\psi(\theta)\wedge\partial_{\theta_2}\psi(\theta)=\sin(\theta_1)\begin{pmatrix}\sin(\theta_1)\cos(\theta_2)\\ \sin(\theta_1)\sin(\theta_2)\\ \cos(\theta_1)\end{pmatrix}=\sin(\theta_1)\ (\partial\varphi)_\psi(\theta)\in T^\perp(S)$$

This implies that

$$T_x(S)=\mathrm{Vect}\left((\partial_{\theta_1}\psi)_{\phi(x)},(\partial_{\theta_2}\psi)_{\phi(x)}\right)\quad\text{and}\quad T_x^\perp(S)=\mathrm{Vect}\left((\partial\varphi)(x)\right)$$

The Riemannian metric on $S_\psi=\mathbb{R}^2$) is given by

$$\begin{aligned}g_{1,1}(\theta)&=\langle(\partial_{\theta_1}\psi)(\theta),(\partial_{\theta_1}\psi)(\theta)\rangle=1\\ g_{2,2}(\theta)&=\langle(\partial_{\theta_2}\psi)(\theta),(\partial_{\theta_2}\psi)(\theta)\rangle=\sin^2(\theta_1)\\ g_{1,2}(\theta)&=g_{2,1}(\theta)=\langle(\partial_{\theta_1}\psi)(\theta),(\partial_{\theta_2}\psi)(\theta)\rangle=0\end{aligned}$$

Up to the circle $S\cap\{x_1=x_2=0\}$ (i.e. for $\theta_1\in\{0,\pi\}$), we have

$$g^{-1}(\theta)=\begin{pmatrix}1&0\\ 0&\sin^{-2}(\theta_1)\end{pmatrix}$$

Notice that

$$\partial_{\theta_1,\theta_2}\psi(\theta)=\begin{pmatrix}-\cos(\theta_1)\sin(\theta_2)\\ \cos(\theta_1)\cos(\theta_2)\\ 0\end{pmatrix}=\tfrac{\cos(\theta_1)}{\sin(\theta_1)}\times\partial_{\theta_2}\psi(\theta)$$

$$\Rightarrow\quad\Gamma^1_{1,2}=0=\Gamma^1_{2,1}\quad\text{and}\quad\Gamma^2_{1,2}(\theta)=\Gamma^2_{2,1}(\theta)=\tfrac{\cos(\theta_1)}{\sin(\theta_1)}$$

In much the same way, we have

$$\begin{aligned}\partial_{\theta_1,\theta_1}\psi(\theta)&=\begin{pmatrix}-\sin(\theta_1)\cos(\theta_2)\\ -\sin(\theta_1)\sin(\theta_2)\\ -\cos(\theta_1)\end{pmatrix}=-(\partial\varphi)_\psi(\theta)\quad\Rightarrow\quad\forall k\in\{1,2\}\ \ \Gamma^k_{1,1}=0\\ \partial_{\theta_2,\theta_2}\psi(\theta)&=\begin{pmatrix}-\sin(\theta_1)\cos(\theta_2)\\ -\sin(\theta_1)\sin(\theta_2)\\ 0\end{pmatrix}=-\sin(\theta_1)\begin{pmatrix}\cos(\theta_2)\\ \sin(\theta_2)\\ 0\end{pmatrix}\end{aligned}$$

In addition, it is readily checked that

$$\partial_{\theta_2,\theta_2}\psi(\theta)\ \perp\partial_{\theta_2}\psi(\theta)\Longleftrightarrow\langle\partial_{\theta_2,\theta_2}\psi(\theta),\partial_{\theta_2}\psi(\theta)\rangle=0\ \Rightarrow\ \Gamma^2_{2,2}=0$$

and

$$\Gamma^1_{2,2}(\theta)=\langle\partial_{\theta_2,\theta_2}\psi(\theta),\partial_{\theta_1}\psi(\theta)\rangle=-\sin(\theta_1)\cos(\theta_1)$$

Using (6.68), we conclude that

$$\begin{aligned}(\Delta\phi^1)_\psi(\theta)&=-\sum_{1\le i,j\le2}\Gamma^1_{i,j}(\theta)\ g^{i,j}(\theta)=-\frac{\Gamma^1_{2,2}(\theta)}{\sin^2(\theta_1)}=\frac{\cos(\theta_1)}{\sin(\theta_1)}=\cot(\theta_1)\\ (\Delta\phi^2)_\psi(\theta)&=0\end{aligned}\tag{6.86}$$

and by (6.40) we have

$$\begin{aligned}(\nabla\phi^1)_\psi(\theta)&=\sum_{1\le j\le2}g^{1,j}(\theta)\ \left(\partial_{\theta_j}\psi\right)(\theta)=(\partial_{\theta_1}\psi)(\theta)\\ (\nabla\phi^2)_\psi(\theta)&=\sum_{1\le j\le2}g^{2,j}(\theta)\ \left(\partial_{\theta_j}\psi\right)(\theta)=\frac{1}{\sin^2(\theta_1)}\ (\partial_{\theta_2}\psi)(\theta)=\frac{1}{\sin(\theta_1)}\begin{pmatrix}-\sin(\theta_2)\\ \cos(\theta_2)\\ 0\end{pmatrix}\end{aligned}\tag{6.87}$$

The geodesics $c_\theta(t) := Exp_\theta(tV) = \begin{pmatrix} c_\theta^1(t) \\ c_\theta^2(t) \end{pmatrix}$, with $V(\theta) \in \mathbb{R}^2$ satisfy the differential equations

$$\begin{cases} \ddot{c}_\theta^{\,1}(t) &= \sin(c_\theta^1(t))\cos(c_\theta^1(t))\ \dot{c}_\theta^{\,2}(t)\ \dot{c}_\theta^{\,2}(t) \\ \ddot{c}_\theta^{\,2}(t) &= -2\,\frac{\cos(c_\theta^1(t))}{\sin(c_\theta^1(t))}\ \dot{c}_\theta^{\,1}(t)\ \dot{c}_\theta^{\,2}(t) \end{cases}$$

with initial conditions

$$c_\theta(0) = \theta \quad \text{and} \quad \dot{c}_\theta(0) = V(\theta)$$

These equations cannot be solved explicitly, and we need to resort to some numerical approximation. The second equation can be rewritten as follows

$$\begin{aligned} \frac{d}{dt}\left(\dot{c}_\theta^{\,2}(t)\ \sin^2(c_\theta^1(t))\right) &= \ddot{c}_\theta^{\,2}(t)\ \sin^2(c_\theta^1(t)) + 2\sin(c_\theta^1(t))\cos(c_\theta^1(t))\ \dot{c}_\theta^{\,1}(t)\ \dot{c}_\theta^{\,2}(t) \\ &= -2\,\frac{\cos(c_\theta^1(t))}{\sin(c_\theta^1(t))}\ \sin^2(c_\theta^1(t))\ \dot{c}_\theta^{\,1}(t)\ \dot{c}_\theta^{\,2}(t) + +2\sin(c_\theta^1(t))\cos(c_\theta^1(t))\ \dot{c}_\theta^{\,1}(t)\ \dot{c}_\theta^{\,2}(t) \\ &= 0 \end{aligned}$$

This shows that

$$\dot{c}_\theta^{\,2}(t)\ \sin^2(c_\theta^1(t)) = \dot{c}_\theta^{\,2}(0)\ \sin^2(c_\theta^1(0))$$

The geodesics $C_x(t) := Exp_x(tW)$ have a more explicit description given by the equations

$$Exp_x(tW) = \cos\left(t\|W(x)\|\right)\ x + \sin\left(t\|W(x)\|\right)\ \frac{W(x)}{\|W(x)\|}$$

We readily check that $C_x(t)$ satisfies the required conditions

$$\dot{C}_x(t) = \left[-\sin\left(t\|W(x)\|\right)\ x + \cos\left(t\|W(x)\|\right)\ \frac{W(x)}{\|W(x)\|)}\right]\ \|W(x)\| \stackrel{t=0}{=} W(x)$$

and

$$C_x(t) \in \mathbb{S}^2 \Longrightarrow \ddot{C}_x(t) = -\|W(x)\|^2\ C_x(t) \in T^\perp(\mathbb{S}^2) \Rightarrow \pi\left(\ddot{C}_x(t)\right) = 0$$

6.7.3 The Torus

The Torus $\mathcal{T}$ can be seen as a surface of revolution obtained by revolving a circle

$$\mathcal{C}(R,r) = \left\{ \begin{pmatrix} R \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} r\cos(\theta_1) \\ 0 \\ r\sin(\theta_1) \end{pmatrix}\ :\ \theta_1 \in \mathbb{R} \right\}$$

of radius r and center $x_1 = R > r$ about the symmetry x_3-axis. The Cartesian coordinates of the Torus are parametrized by the function

$$\psi\ :\ \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \in \mathbb{R}^2 \mapsto \psi(\theta) = \begin{pmatrix} (R + r\cos(\theta_1))\cos(\theta_2) \\ (R + r\cos(\theta_1))\sin(\theta_2) \\ r\sin(\theta_1) \end{pmatrix}$$

Alternatively, $\mathcal{T} = \varphi^{-1}(0)$ can be represented as the null level set of the function

$$\varphi\ :\ x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mapsto \varphi(x) = \left(R - \sqrt{x_1^2 + x_2^2}\right)^2 + x_3^2 - r^2$$

After some elementary manipulations, we find that

$$\partial_{\theta_1}\psi(\theta) = \begin{pmatrix} -r\sin(\theta_1)\cos(\theta_2) \\ -r\sin(\theta_1)\sin(\theta_2) \\ r\cos(\theta_1) \end{pmatrix} \quad \text{and} \quad \partial_{\theta_2}\psi(\theta) = \begin{pmatrix} -(R+r\cos(\theta_1))\sin(\theta_2) \\ (R+r\cos(\theta_1))\cos(\theta_2) \\ 0 \end{pmatrix}$$

and

$$\frac{R-\sqrt{x_1^2+x_2^2}}{\sqrt{x_1^2+x_2^2}}\, x_1 \overset{x=\psi(\theta)}{=} \frac{-r\cos(\theta_1)}{(R+r\cos(\theta_1))}\,(R+r\cos(\theta_1))\cos(\theta_2) = -r\cos(\theta_1)\cos(\theta_2)$$

$$\Longrightarrow \partial\varphi(x) = 2\begin{pmatrix} -x_1 \dfrac{R-\sqrt{x_1^2+x_2^2}}{\sqrt{x_1^2+x_2^2}} \\ -x_2 \dfrac{R-\sqrt{x_1^2+x_2^2}}{\sqrt{x_1^2+x_2^2}} \\ x_3 \end{pmatrix} \overset{x=\psi(\theta)}{=} 2r\begin{pmatrix} \cos(\theta_1)\cos(\theta_2) \\ \cos(\theta_1)\sin(\theta_2) \\ \sin(\theta_1) \end{pmatrix}$$

In addition, we have

$$\begin{array}{rcl} \partial_{\theta_1}\psi(\theta) & \perp & \partial_{\theta_2}\psi(\theta) \quad (\Leftrightarrow \langle \partial_{\theta_1}\psi(\theta), \partial_{\theta_2}\psi(\theta)\rangle = 0) \\ \|\partial_{\theta_1}\psi(\theta)\| & = & r^2 \quad \text{and} \quad \|\partial_{\theta_2}\psi(\theta)\| = (R+r\cos(\theta_1))^2 \end{array}$$

and

$$\begin{aligned} \partial_{\theta_2}\psi(\theta)\wedge\partial_{\theta_1}\psi(\theta) &= \begin{pmatrix} -(R+r\cos(\theta_1))\sin(\theta_2) \\ (R+r\cos(\theta_1))\cos(\theta_2) \\ 0 \end{pmatrix} \wedge \begin{pmatrix} -r\sin(\theta_1)\cos(\theta_2) \\ -r\sin(\theta_1)\sin(\theta_2) \\ r\cos(\theta_1) \end{pmatrix} \\ &= r(R+r\cos(\theta_1))\begin{pmatrix} \cos(\theta_1)\cos(\theta_2) \\ \cos(\theta_1)\sin(\theta_2) \\ \sin(\theta_1) \end{pmatrix} \end{aligned}$$

This shows that

$$g(\theta) = \begin{pmatrix} r^2 & 0 \\ 0 & (R+r\cos(\theta_1))^2 \end{pmatrix} \quad \text{and} \quad g^{-1}(\theta) = \begin{pmatrix} r^{-2} & 0 \\ 0 & (R+r\cos(\theta_1))^{-2} \end{pmatrix}$$

$$\Longrightarrow \sqrt{\det(g(\theta))} = r(R+r\cos(\theta_1))$$

Using (6.40) we have

$$\begin{aligned} (\nabla\phi^1)_\psi(\theta) &= r^{-2}\,(\partial_{\theta_1}\psi)(\theta) \\ (\nabla\phi^2)_\psi(\theta) &= (R+r\cos(\theta_1))^{-2}\,(\partial_{\theta_2}\psi)(\theta) \end{aligned}$$

and

$$\begin{array}{rcl} \big\langle (\nabla\phi^1)_\psi(\theta), (\nabla\phi^2)_\psi(\theta)\big\rangle & = & 0 \\ \big\langle (\nabla\phi^1)_\psi(\theta), (\nabla\phi^1)_\psi(\theta)\big\rangle & = & r^{-2} \quad \text{and} \quad \big\langle (\nabla\phi^2)_\psi(\theta), (\nabla\phi^2)_\psi(\theta)\big\rangle = (R+r\cos(\theta_1))^{-2} \end{array}$$

By (6.65), we also find that

$$\begin{aligned} (\Delta\phi^1)_\psi(\theta) &= \frac{1}{r(R+r\cos(\theta_1))}\,\partial_{\theta_1}\left(r(R+r\cos(\theta_1))\,r^{-2}\right) = -\frac{\sin(\theta_1)}{r(R+r\cos(\theta_1))} \\ (\Delta\phi^2)_\psi(\theta) &= \frac{1}{r(R+r\cos(\theta_1))}\,\partial_{\theta_2}\left(r(R+r\cos(\theta_1))\,(R+r\cos(\theta_1))^{-2}\right) = 0 \end{aligned}$$

Chapter 7

Markov Chain Monte Carlo models

7.1 Boltzmann-Gibbs target measures

Markov chain Monte Carlo algorithms are rather standard stochastic simulation methods for sampling from a given target distribution, say π on some state space E. The prototype of target measure is given by Boltzmann-Gibbs measures of the following form

$$\pi(dx) = \Psi_G(\lambda)(dx) = \frac{1}{\lambda(G)}\ G(x)\ \lambda(dx) \tag{7.1}$$

where λ stands for some reference measure and G some potential function on some state space E.

Boltzmann-Gibbs measures are one of the most central mathematical models of classical statistical physics. In this context, the main problem is to deduce macroscopic equilibrium behaviors of thermodynamic physical systems, from complex disordered microscopic interacting structures.

As their name indicates, Boltzmann-Gibbs measures were introduced independently in the early 1900s by Ludwig Boltzmann and Josiah Willard Gibbs in their seminal studies on statistical entropy theory and micro-macro canonical ensemble theory [60, 287]. In the late 1960s, Roland Lvovich Dobrushin [231] and Oscar E. Lanford and David Ruelle [396] and also developed a new theory to design probability measures on finite product spaces, by specification systems of conditional distributions w.r.t. the complement of finite volume measures, with prescribed boundary conditions. The two prototypes of physical systems are particles in liquid-vapor models of real gases, interacting via Van der Waals forces, and atoms' configurations and their magnetic moments in crystal lattices of ferromagnetic metals (iron, colbalt, nickel) in thermal equilibrium. In this context, adjacent atoms and their electronic configurations tend to have the same angular moment (i.e., the same spins).

The central idea behind MCMC methodologies is to design a judicious Markov transition $M(x, dy)$, with nice stability properties, that has the target probability measure $\pi = \pi M$ as its invariant measure. After a rather large number of runs, and when the chain is sufficiently stable, the ergodic theorem tells us that the occupation measures of the random states X_n of the chain with Markov transition M approximate π.

In computational physics as well as in stochastic optimization we are interested in computing Boltzmann-Gibbs distributions associated with an inverse temperature parameter β. These distributions have the following form

$$\mu_\beta(dx) = \frac{1}{\mathcal{Z}_\beta}\, e^{-\beta V(x)}\, \lambda(dx) \tag{7.2}$$

where λ stands for some distribution on some state space S, and V some function on E. The parameter β is interpreted as an inverse temperature parameter. It is also often important to compute the normalizing constants (a.k.a. partition functions in physics) $\mathcal{Z}_\beta = \lambda\left(e^{-\beta V}\right)$

Let us present an example of Boltzmann-Gibbs measure arising in Operation Research. We are given a finite state space $\mathcal{E}_m = \{e_1, \ldots, e_m\}$ equipped with some metric d. We can think of a a finite number of cities and the distance between them. One typical problem known as the traveling salesman problem, consists in finding a way to visit all the cities with a minimal travel distance. We can model a given sequence of visits by a permutation σ on the index set $\{1, \ldots, m\}$. In this situation, the state space E is given by the set of these permutations equipped with the uniform probability measure $\lambda(\sigma) = 1/m!$, and the energy function V is defined by

$$V(\sigma) = \sum_{p=1}^{m} d(e_{\sigma(p)}, e_{\sigma(p+1)})$$

It is not difficult to check that

$$\lim_{\beta\to\infty} \mu_\beta(\sigma) = \mu_\infty(\sigma) := \frac{1}{\mathrm{Card}(V^\star)}\, 1_{V^\star}(\sigma) \tag{7.3}$$

where $\mathrm{Card}(V^\star)$ stands for the cardinal of the set $V^\star = \{\sigma \in E \ : \ V(\sigma) = \inf_E V\}$ of the optimal traveling strategies. Indeed, we have

$$\sum_{\sigma\in E} e^{-\beta[V(\sigma)-V_\star]} = \mathrm{Card}\,(\mathcal{V}_\star) + \sum_{\sigma\notin\mathcal{V}_\star} e^{-\beta[V(\sigma)-V_\star]} \to_{n\uparrow\infty} \mathrm{Card}\,(\mathcal{V}_\star)$$

This implies that μ_β converges to the uniform measure on the set $\mathcal{V}_\star$; that is, we have that

$$\mu_\beta(\sigma) = \frac{e^{-\beta[V(\beta)-V_\star]}}{\sum_{\tau\in E} e^{-\beta[V(\tau)-V_\star]}} \to_{\beta\uparrow\infty} \mu_\infty(\sigma) := \frac{1}{\mathrm{Card}\,(\mathcal{V}_\star)}\, 1_{\mathcal{V}_\star}(\sigma)$$

This result shows that at low temperature (i.e. $\beta \uparrow \infty$), the sampling of the distribution μ_β amounts of choosing randomly an unknown optimal solution of the problem.

Boltzmann-Gibbs measures on manifolds arise in global optimization problems with nonlinear constraints. In this situation, the state space E is given by the q-dimensional null level set $S = \varphi^{-1}(0)$ of some smooth function $\varphi \ : \ x \in \mathbb{R}^{r=p+q} \mapsto \mathbb{R}^q$ s.t. $\mathrm{rank}(\partial\varphi(x)) = q$, for any $x \in \mathbb{R}^r$. The corresponding optimization problem is often expressed in the form

$$\arg\min\{V(x) \ : \ x \in S\} = \arg\min\{V(x) \ : \ x \in \mathbb{R}^r \text{ s.t. } \varphi(x) = 0\}$$

The manifold S can be equipped with a Riemannian structure associated with some chart $\phi \ : \ x \in S \mapsto \phi(x) \in S_\phi$ (and some parametric description $\psi = \phi^{-1} \ : \ \theta \in S_\phi \mapsto S$).

In Bayesian inference, Boltzmann-Gibbs measures represent the posterior distribution of some unknown random parameter $\Theta = \theta$ given some partial and noisy observations $Y = y$. In this context, λ represents the prior distribution of Θ, and V the log-likelihood function of the parameter w.r.t. a given observation; that is, for $\beta = 1$ we have that

$$\lambda(d\theta) = \mathbb{P}(\Theta \in d\theta) \quad \text{and} \quad V(\theta) = -\log P_\theta(y)$$

where $P_\theta(y) = p(y|\theta)$ stands for the conditional density of $Y = y$ w.r.t. some reference measure $\lambda'(dy)$ on the observation space. In this situation, we have the Bayes' rule

$$\mu_{\beta=1}(d\theta) = \frac{1}{\mathcal{Z}_1}\, e^{-V(\theta)}\, \lambda(d\theta) \propto p(y|\theta)\, \mathbb{P}(\Theta \in d\theta) \propto \mathbb{P}(\Theta \in d\theta \mid Y = y)$$

Directed polymer models arising in statistical physics are defined in much the same way. For instance, the micro-state of a system consists of d particles $x_i = (p_i, r_i)$, with a momentum vector p_i and a position coordinate $r_i = (r_i^1, r_i^2, r_i^3)$, with $1 \leq i \leq d$. The energy of the system is given by some function

$$V(x) := \sum_{i=1}^{d} \left(\frac{1}{2m}\, \|p_i\|^2 + mgr_i^1 \right)$$

where m represents the mass of the particle, r_i^1 its height, and g the gravitation constant. The probability distribution of the physical system at inverse temperature β_n is again given by the Boltzmann-Gibbs measures (7.2), with the Lebesgue measure λ. For a more thorough discussion on these models, we refer the reader to [149, 213, 216], and references therein.

More generally, the r.v. X may represent the random paths $X = (X_0, \ldots, X_n)$ of a given reference Markov process X_n, taking values in some state spaces E_n. In this situation, the reference measure λ in (7.1) is often given by the distribution of the random path

$$\lambda(dx) = \lambda(d(x_0, \ldots, x_n)) = \mathbb{P}((X_0, \ldots, X_n) \in d(x_0, \ldots, x_n))$$

where $d(x_0, \ldots, x_n)$ stands for an infinitesimal neighborhood f the path $x = (x_0, \ldots, x_n) \in S := (E_0 \times \ldots \times E_n)$. The potential weight functions $G(x)$ are often given by a product formula

$$G(x) = G(x_0, \ldots, x_n) = \prod_{0 \leq k \leq n} G_k(x_k)$$

For instance, for indicator functions $G_k = 1_{A_k}$ of some measurable subset $A_k \subset E_k$, we have

$$\begin{aligned} \pi(dx) = \frac{1}{\lambda(G)}\, G(x)\, \lambda(dx) \quad &\propto \quad 1_{A_0 \times \ldots \times A_n}(x_0, \ldots, x_n)\, \mathbb{P}((X_0, \ldots, X_n) \in d(x_0, \ldots, x_n)) \\ &\propto \quad \mathbb{P}((X_0, \ldots, X_n) \in d(x_0, \ldots, x_n) \mid (X_0, \ldots, X_n) \in (A_0 \times \ldots \times A_n)) \end{aligned}$$

In probability theory and computational physics, these Boltzmann-Gibbs measures associated with some Markov chain process are also termed Feynman-Kac measures. Their analysis and their particle interpretations are discussed in section 9.1.4.

7.2 Metropolis and Hasting models

The Metropolis-Hastings algorithm is the most famous MCMC model of current use in practice.

Firstly, we choose a Markov transition K to explore randomly the state S. We further assume that $K(x, dy)$ and the target measure $\pi(dy)$ have a density with respect to some reference measure $\lambda(dy)$; that is, we have that

$$K(x, dy) = k(x, y)\ \lambda(dy) \quad \text{and} \quad \pi(dy) = h(y)\ \lambda(dy)$$

with some density functions $k(x, y)$ and $h(y)$ s.t.

$$h(y)k(y, x) = 0 \implies h(x)k(x, y) = 0$$

We set

$$G(x, y) := \frac{h(y)k(y, x)}{h(x)k(x, y)}$$

with the convention $0/0 = 0$. For more general models, we take

$$G(x, y) = \frac{\pi(dy)K(y, dx)}{\pi(dx)K(x, dy)} \tag{7.4}$$

For Boltzmann-Gibbs measures π of the form (7.1), it is readily checked that the function G doesn't depend on the normalizing constant $\lambda(G)$, and it is given by the formula

$$G(x, y) = \frac{G(y)}{G(x)} \times \frac{\lambda(dy)K(y, dx)}{\lambda(dx)K(x, dy)}$$

In addition, when the proposal transition K is reversible w.r.t. the measure λ, the function G takes the simpler form

$$G(x, y) = G(y)/G(x) \tag{7.5}$$

The Metropolis-Hastings model is a Markov chain with μ-reversible acceptance-rejection style transitions of the following form

$$M(x, dy) = K(x, dy)\ a(x, y) + (1 - \int K(x, dz)\ a(x, z))\ \delta_x(dy) \tag{7.6}$$

To guarantee the reversibility property, we often chose one of the following acceptance rates

$$a = G/(1 + G) \qquad \text{or} \qquad a = 1 \wedge G \tag{7.7}$$

When the proposal transition $K(x, .) = \nu$ is given by some probability measure ν, that doesn't depends on the current state x, the resulting MCMC sampler is sometimes called an independent Metropolis Hastings model.

The Metropolis-Hasting transition (7.6) associated with one of the acceptance rates a given in (7.7) is reversible w.r.t. the target measure π; that is we have that

$$\pi(dx)M(x, dy) = \pi(dy)M(y, dx)$$

When $a = 1 \wedge G$, for any $x \neq y$ we have

$$\begin{aligned} \pi(dx)M(x, dy) &= \pi(dx)K(x, dy)\ a(x, y) \\ &= \lambda(dx)\ h(x)\ k(x, y)\lambda(dy)\ \left\{1 \wedge \frac{(h(y)k(y, x))}{(h(x)k(x, y))}\right\} \\ &= \lambda(dx)\lambda(dy)\ \ \{(h(x)k(x, y)) \wedge (h(y)k(y, x))\} \end{aligned}$$

This formula is clearly symmetric w.r.t. x and y.

When $a = G/(1+G)$, for any $x \neq y$ we have

$$a(x,y) = \frac{\frac{h(y)k(y,x)}{h(x)k(x,y)}}{1+\frac{h(y)k(y,x)}{h(x)k(x,y)}} = \frac{h(y)k(y,x)}{h(x)k(x,y)+h(y)k(y,x)}$$

and

$$\begin{aligned}\pi(dx)M(x,dy) &= \pi(dx)K(x,dy)\ a(x,y)\\ &= \lambda(dx)\lambda(dy)\ \frac{(h(y)k(y,x))(h(x)k(x,y))}{h(x)k(x,y)+h(y)k(y,x)}\end{aligned}$$

This formula is again symmetric w.r.t. x and y.

■

For a detailed discussion on this model, we refer the reader to the pioneering article by N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller [454], the more recent review article by N. Metropolis [453], and the series of articles by P. Diaconis [223, 224, 226].

The mathematical analysis of this Markov chain model is also well developed. We refer the reader to the series of seminal articles by P. Diaconis and his co-authors [218, 219, 220, 221, 225, 227]. These works reveals fascinating connexions between the design and the performance analysis of MCMC model with powerful pure and applied mathematical techniques, ranging from representation theory, micro-local analysis, log-Sobolev inequalities, and spectral analysis.

Besides the fact that these techniques provide very sharp rates of convergence, it is clearly of course out of the scope of this book to review these methods. In the end of this section, we content ourselves with presenting one of the simplest way to analyze the convergence of an MCMC algorithm.

Suppose that K satisfies the minorization condition

$$K^m(x,dy) \geq \epsilon\ \nu(dy)$$

for some interger $m \geq 1$, some $\epsilon \in]0,1]$ and some probability measure ν. In this situation, we clearly have that

$$M^m(x,dy) \geq \epsilon'\ \nu(dy)$$

with

$$\epsilon' = \epsilon \inf_{x_0 \rightsquigarrow \ldots \rightsquigarrow x_m} \left[\prod_{0 \leq p < m} a(x_p, x_{p+1}) \right]$$

The r.h.s. infimum is taken over all sequences $(x_0, \ldots, x_m)$ of states in S of length m and s.t.

$$k(x_p, x_{p+1}) > 0$$

Whenever $\epsilon > 0$, the Dobrushin contraction coefficient of M^m is s.t. $\beta(M^m) < 1$. For a more detailed discussion on MCMC models, and their stochastic analysis we refer the reader to the review articles by P. Diaconis [217, 222, 227], and references therein.

Our next objective is to sample a centered Gaussian random variable Z with unit variance restricted to some set A, say $A =]a,b[$, for some $-\infty \leq a < b \leq \infty$. We let λ be the distribution of Z, and we set

$$\pi(dz) := \frac{1}{\lambda(A)}\ 1_A(x)\ \lambda(dx) = \mathbb{P}\left(Z \in dz \mid Z \in A\right)$$

Of course, we can use the distribution function $F(z) = \mathbb{P}(Z \leq z)$ of the Gaussian r.v. and set

$$Z_{a,b} = F^{-1}\left(F(a) + U\ (F(b) - F(a))\right) \tag{7.8}$$

It is an elementary exercise to check that $\text{Law}(Z_{a,b}) = \pi$. Nevertheless, the function F requires to integrate the Gaussian density from $-\infty$ up to any state z, using some kind of numerical approximation scheme.

Another strategy is to use a rejection simulation technique. In this case, we sample a sequence of independent copies of Z and we accept the ones that hit the desired set A.

Next, we describe an alternative approach based on Markov chain simulation.

We let Z_n be a sequence of independent copies of Z. Next, we design a Markov chain X_n with invariant measure π.

Suppose, the chain $X_n \in A$ at some time step $n \geq 0$. Starting from this point, we set

$$Y_{n+1} = \sqrt{1-\epsilon_n}\ X_n + \sqrt{\epsilon_n}\ Z_n$$

If Y_{n+1} we accept the move and we set X_{n+1}. Otherwise, we stay in the same place $X_{n+1} = X_n$. The Markov transition of the chain $X_n \rightsquigarrow X_{n+1}$ is given by

$$M(x, dy) = K(x, dy)\ 1_A(x) + (1 - K(1_A)(x))\ \delta_x(dy)$$

We prove that $\pi M = \pi$ as follows. Recalling that the transition $K(x, dy)$ is a reversible w.r.t. to the Gaussian distribution λ, for any bounded functions f_1, f_2 on $\mathbb{R}$ we have that

$$\begin{aligned}
\pi M(f) &\propto \lambda(1_A M(f)) \\
&= \lambda(1_A K(1_A\ f)) + \lambda(1_A(1 - K(1_A)\ f_2) \\
&= \lambda(K(1_A)1_A\ f_2) + \lambda(1_A f) - \lambda(1_A\ K(1_A)\ f_2) = \lambda(1_A f) \propto \pi(f)
\end{aligned}$$

This clearly implies that $\pi = \pi M$.

7.3 Gibbs-Glauber dynamics

We let π be some target measure defined on some product state space $S = (S_1 \times S_2)$. We assume that the following disintegration property is satisfied

$$\pi(d(x_1, x_2)) = \pi_1(dx_1)L_{1,2}(x_1, dx_2) = \pi_2(dx_2)L_{2,1}(x_2, dx_1)$$

with the first, and second marginals, π_1 and π_2, and the corresponding conditional probability measures $L_{1,2}$ and $L_{2,1}$.
The Gibbs sampler is the Markov chain with the elementary transition

$$M = K_1 K_2$$

with the transitions K_i given for any $i \in \{1, 2\}$ by

$$\begin{aligned}
K_1((x_1, x_2), d(y_1, y_2)) &:= \delta_{x_1}(dy_1)L_{1,2}(y_1, dy_2) \\
K_2((x_1, x_2), d(y_1, y_2)) &:= \delta_{x_2}(dy_2)L_{2,1}(y_2, dy_1)
\end{aligned}$$

We can alternatively choose the Markov transitions

$$M = K_2 K_1 \quad \text{or} \quad M = \frac{1}{2}\ K_1 + \frac{1}{2}\ K_2$$

The transitions K_1 and K_2 are reversible w.r.t. the measure π. In addition, the Metropolis Hasting transitions M_1, and resp. M_2, with proposal transition K_1, and resp. K_2, and acceptance rate $a = 1 \wedge G$ with G given by (7.4) has unit acceptance rate.

Proof :
We check this claim using the fact that

$$\begin{aligned}
&\pi(d(y_1, y_2)) \times K_1((y_1, y_2), d(x_1, x_2)) \\
&= \pi_1(dy_1) L_{1,2}(y_1, dy_2) \times \delta_{y_1}(dx_1) L_{1,2}(x_1, dx_2) \\
&= \underbrace{\pi_1(dy_1)\delta_{y_1}(dx_1)}_{=\pi_1(dx_1)\delta_{x_1}(dy_1)} \times (L_{1,2}(y_1, dy_2) L_{1,2}(x_1, dx_2))
\end{aligned}$$

This formula is clearly symmetric w.r.t. $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Thus, we have

$$G((x_1, x_2), (y_1, y_2)) = \frac{\pi(d(y_1, y_2)) \times K_1((y_1, y_2), d(x_1, x_2))}{\pi(d(x_1, x_2)) \times K_1((x_1, x_2), d(y_1, y_2))} = 1$$

■

The resulting Markov chain model is often called the Gibbs sampler or the Glauber dynamics.

Example 7.3.1 *Suppose we want to select uniformly a point $Z = (X, Y)$ in the unit disk*

$$D := \{(x, y) \in [-1, 1]^2 \ : \ x^2 + y^2 \leq 1\}$$

Here again, we can use a rejection simulation technique. We sample a sequence of independent random variables on $[-1, 1]^2$ and we accept the ones that hit the desired set D. Notice that the invariant measure π on D is defined by

$$\begin{aligned}
\pi(d(x, y)) &\propto 1_D(x, y)\ dxdy \\
&= 1_{[-1,1]}(x)\ dx\ 1_{[-\sqrt{1-x^2}, +\sqrt{1-x^2}]}(y)\ dy \\
&= 1_{[-1,1]}(y)\ dy\ 1_{[-\sqrt{1-y^2}, +\sqrt{1-y^2}]}(x)\ dx
\end{aligned} \tag{7.9}$$

This implies that

$$\begin{aligned}
\mathbb{P}\left(Y \in dy \mid X = x\right) &\propto 1_{[-\sqrt{1-x^2}, +\sqrt{1-x^2}]}(y)\ dy \\
\mathbb{P}\left(X \in dx \mid Y = y\right) &\propto 1_{[-\sqrt{1-y^2}, +\sqrt{1-y^2}]}(x)\ dx
\end{aligned}$$

Pick any initial point (X_0, Y_0) in D. The next state of the chain (X_1, Y_1) is defined as follows. Firstly, we choose uniformly a point X_1 on $\left[-\sqrt{1 - Y_0^2}, +\sqrt{1 - Y_0^2}\right]$. Then, we choose uniformly a state Y_1 on $\left[-\sqrt{1 - X_1^2}, +\sqrt{1 - X_1^2}\right]$. Iterating these couple of transitions, we construct a Markov chain evolving inside the the unit disk D

$$Z_n := \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \rightsquigarrow \begin{pmatrix} X_{n+1} \\ Y_n \end{pmatrix} \rightsquigarrow Z_{n+1} = \begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} \tag{7.10}$$

After some elementary computations, we prove that π is the invariant measure of the chain Z_n.

These constructions can be extended to product state spaces of any dimension. More formally, we let π be some target measure on some state space $S = E^I$, where I stands for some finite set.

7.4 The Propp and Wilson sampler

In the further development of this section $M(x,y)$ stands for an aperiodic and irreducible Markov transition on some finite set S.

A random mapping F is said to be M-compatible as soon as we have for any $(x,y) \in S^2$

$$\mathbb{P}\left(F(x)=y\right)=M(x,y)$$

The existence of M-compatible mappings is proved as follows:

Up to some change of label, there is no loss of generality to assume that the state space $S=\{1,\ldots,d\}$, with $d=\mathrm{Card}(S)$. In this notation, a mapping F is characterized by a column random vector $F=(F(1),\ldots,F(d))'$

We let $(U_i)_{1\leq i\leq d}$ be a sequence of independent and uniform r.v. on $[0,1[$, and we set

$$F(i)=\sum_{1\leq j\leq d} j\ 1_{[\sum_{1\leq k<j} M(i,k),\ \sum_{1\leq k\leq j} M(i,k)[}(U_i) \tag{7.11}$$

By construction, the random states $(F(i))_{1\leq i\leq d}$ are *independent r.v.* and we have

$$\mathbb{P}\left(F(i)=j\right)=\mathbb{P}\left(\sum_{1\leq k<j} M(i,k)\leq U_i\leq \sum_{1\leq k\leq j} M(i,k)\right)=M(i,j)$$

From the above construction, we notice that F is not necessarily a one to one mapping.

In this notation, the Markov chain with elementary transition M is defined for any $n\geq 0$ by the recursion

$$X_{n+1}=F_n(X_n)=F_n(F_{n-1}(X_{n-1}))=\ldots=\left(F_n\circ\ldots\circ F_1\circ F_0\right)(X_0)$$

where F_n, with $n\in\mathbb{N}$, stands for a sequence of independent copies of the mapping F.

Given a sequence of independent copies $(F_n)_{n\geq 0}$ of the mapping F, we let

$$\overleftarrow{\boldsymbol{F_n}}:=F_0\circ F_1\circ\ldots\circ F_n\overset{law}{=}F_n\circ\ldots\circ F_1\circ F_0:=\overrightarrow{\boldsymbol{F_n}}$$

We also let $\overleftarrow{T}$ and $\overrightarrow{T}$ be the forward and backward coalescent times

$$\begin{aligned}\overleftarrow{T} &= \inf\left\{n\ :\ \mathrm{Card}\left(\ \overleftarrow{\boldsymbol{F_n}}(S)\ \right)=1\right\}\\ \overrightarrow{T} &= \inf\left\{n\ :\ \mathrm{Card}\left(\ \overrightarrow{\boldsymbol{F_n}}(S)\ \right)=1\right\}\overset{law}{=}\overleftarrow{T}\end{aligned}$$

The backward mapping $\overleftarrow{\boldsymbol{F_n}}$ is better interpreted as running the chain forward from some random state X_{-n} up to the state X_1

$$X_1=F_0(X_0)=F_0(F_{-1}(X_{-1}))=\ldots=\left(F_0\circ F_{-1}\circ\ldots\circ F_{-n}\right)(X_{-n}) \tag{7.12}$$

where F_n, with $n\in\mathbb{Z}$, stands for independent copies of F. In this situation, the initial condition is X_{-n} and X_1 is the terminal state of the chain after $(n+1)$ forward interactions.

We further assume that the M-compatible mapping F is chosen so that

$$\mathbb{P}\left(\overleftarrow{T} < \infty\right) = 1 = \mathbb{P}\left(\overrightarrow{T} < \infty\right) \tag{7.13}$$

In this situation, the value of $\overleftarrow{\boldsymbol{F}_{\overleftarrow{\boldsymbol{T}}}}(x) := Y$ doesn't depend on the state variable x and it is distributed according to the invariant measure of the chain $\pi = \pi M$.

Proof :
By construction the value $\overleftarrow{\boldsymbol{F}_{\overleftarrow{\boldsymbol{T}}}}$, and a fortiori the one of $\overleftarrow{\boldsymbol{F}_{\overleftarrow{\boldsymbol{T}}}}(x) := Y$ doesn't depends on the variable x. This implies that

$$\begin{aligned} & T \leq n \\ & \Rightarrow \left(F_0 \circ F_1 \circ \ldots \circ F_n\right)(x) = \left(F_0 \circ F_1 \circ \ldots \circ F_T\right) \circ \left(F_{T+1} \circ \ldots \circ F_n(x)\right) = Y \end{aligned}$$

We conclude that

$$\begin{aligned} \mathbb{P}(Y = y) \overset{\infty \leftarrow n}{\longleftarrow} & \quad \mathbb{P}\left(\left(F_0 \circ F_1 \circ \ldots \circ F_n\right)(x) = y\right) \\ = & \quad \mathbb{P}\left(\left(F_n \circ \ldots \circ F_0\right)(x) = y\right) \overset{n \to \infty}{\longrightarrow} \pi(y) \end{aligned}$$

This ends the proof of the theorem. ∎

The coalescent condition is not satisfied for any M-compatible mappings. For instance when $S = \{1, 2\}$ and $M(i, j) = 1/2$ for any $i, j \in S$ it is readily checked that the mapping F defined by

$$\mathbb{P}\left((F(1), F(2)) = (1, 2)\right) = 1/2 = \mathbb{P}\left((F(1), F(2)) = (2, 1)\right)$$

is M-compatible but the above condition is not met. Indeed, we have that

$$\mathbb{P}(F(1) = 1) = 1/2 = \mathbb{P}(F(1) = 2) \quad \text{and} \quad \mathbb{P}(F(2) = 1) = 1/2 = \mathbb{P}(F(2) = 2)$$

but $\overleftarrow{\boldsymbol{F_n}}$ and $\overrightarrow{\boldsymbol{F_n}}$ are random permutations of the states $\{1, 2\}$.

Nevertheless the mappings defined in (7.11) satisfy the desired condition. To check this claim, we notice that

$$\begin{aligned} \mathbb{P}\left((F(1), F(2)) = (1, 2)\right) &= \mathbb{P}\left((F(1), F(2)) = (1, 1)\right) \\ &= \mathbb{P}\left((F(1), F(2)) = (2, 1)\right) \\ &= \mathbb{P}\left((F(1), F(2)) = (2, 2)\right) = 1/4 \end{aligned}$$

In this situation, we have

$$\mathbb{P}(F(2) = 1) = \mathbb{P}(F(1) = 1) = 1/4 + 1/4 = 1/2 = \mathbb{P}(F(1) = 2) = \mathbb{P}(F(1) = 2)$$

In addition, we have

$$\mathbb{P}\left(\overleftarrow{T} \leq 1\right) \geq \mathbb{P}\left(\text{Card}(F(\{1, 2\})) = 1\right) = 1/2 > 0 \Rightarrow \mathbb{P}\left(\overleftarrow{T} > 1\right) \leq 1/2 < 1$$

Recalling that

$$\mathbb{P}\left(\overleftarrow{T} > n \mid \overleftarrow{T} > (n-1)\right)$$

$$\propto \mathbb{E}\left(\mathbb{P}\left(\mathrm{Card}\left(\overleftarrow{\boldsymbol{F_{n-1}}} \circ F_n(S)\right) > 1 \mid \overleftarrow{\boldsymbol{F_{n-1}}}\right) 1_{\mathrm{Card}\left(\overleftarrow{\boldsymbol{F_{n-1}}}(S)\right)>1}\right)$$

$$= \mathbb{E}\left(\underbrace{\mathbb{P}\left(\mathrm{Card}\left(\overleftarrow{\boldsymbol{F_1}}(S)\right) > 1 \mid \overleftarrow{\boldsymbol{F_0}}\right)_{\overleftarrow{\boldsymbol{F_0}}=\overleftarrow{\boldsymbol{F_{n-1}}}}}_{\leq 1/2} \quad 1_{\mathrm{Card}\left(\overleftarrow{\boldsymbol{F_{n-1}}}(S)\right)>1}\right)$$

This implies that

$$\mathbb{P}\left(\overleftarrow{T} > n\right) = \mathbb{P}\left(\overleftarrow{T} > n \mid \overleftarrow{T} > (n-1)\right) \times \mathbb{P}\left(\overleftarrow{T} > (n-1)\right) \leq 1/2^n$$

from which we conclude that $\mathbb{P}\left(\overleftarrow{T} < \infty\right) = 1$.

Notice that the Propp and Wilson scheme requires to store all the values of the functions F_n. This drawback reflects the main limitation of applying the Propp and Wilson sampler in large state spaces.

Nevertheless, we can overcome this difficulty when the state space S is equipped with a partial order with a minimal and a maximal state, $x_{min} \leq x \leq x_{max}$, for any $x \in S$. In this case, the strategy is to find a judicious monotone M-compatible mapping F. Combining the interpretation (7.12) with the fact that

$$\begin{array}{l} \mathrm{Card}\left(F_0 \circ F_{-1} \circ \ldots \circ F_{-n}\right)(S) = 1 \\ \Leftrightarrow \\ \left(F_0 \circ F_{-1} \circ \ldots \circ F_{-n}\right)(x_{min}) = \left(F_0 \circ F_{-1} \circ \ldots \circ F_{-n}\right)(x_{max}) \end{array}$$

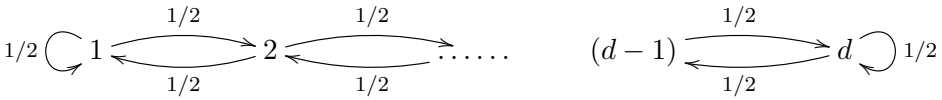
we only need to store the values of two chains starting at x_{min} and x_{max}. This also shows that the coalescence property (7.13) of *monotone* mapping F is granted as soon as the chain is ergodic.

The drawback is that the desired coalescence may not appear after some initially chosen number n steps. In this case, we need to restart the simulation with a larger number of steps. In practice we often choose these numbers of the form 2^k, with $k \geq 1$.

For a more detailed discussion on this simulation technique, we refer the reader to the book of S. Asmussen, P. W. Glynn [20]. The website of D.B. Wilson on perfect sampling with Markov chain also contains a rather complete list of references on this subject.

We end this section with some examples of monotone M-compatible mappings.

Example 7.4.1 (The ladder chain) *We consider the ladder Markov chain X_n defined by the following transition diagram*



We also consider the couple of monotone mappings

$$F_+(x) = \left\{ \begin{array}{rcl} x+1 & \text{for} & x \in \{1, \ldots, d-1\} \\ d & \text{for} & x = d \end{array} \right.$$

and

$$F_-(x) = \left\{ \begin{array}{rcl} 1 & \text{for} & x = 1 \\ x-1 & \text{for} & x \in \{2, \ldots, d-1\} \end{array} \right.$$

Given some uniform r.v. U on $[0,1]$, we set

$$F = 1_{[0,1/2[}(U)\ F_- + 1_{[1/2,1]}(U)\ F_+$$

It is a simple exercise to check that this random mapping is monotone and compatible w.r.t. the Markov transition M of the ladder chain.

7.5 Multilevel annealing type models

7.5.1 Simulated annealing model

We suppose that we are given a sequence of target measures π_n defined in terms of a sequence of Boltzmann-Gibbs measures

$$\pi_n(dx) = \mu_{\beta_n}(dx) = \frac{1}{\mathcal{Z}_{\beta_n}}\ e^{-\beta_n V(x)}\ \lambda(dx) \tag{7.14}$$

associated with some inverse temperature parameter $\beta_n \uparrow \infty$, some non negative potential function V and some reference measure λ on some state space S. Several examples of Boltzmann-Gibbs measures are discussed in section 7.1, including the Ising model and the traveling salesman problem.

For finite state spaces equipped with the counting measure λ, we have seen in (7.3) that these measures converge to the uniform measure on the subset of all global minima of the potential function V, as β_n tends to ∞. This shows that the sampling of these measure at low temperature is equivalent to that of sampling uniformly a state with minimal energy. Since most of the time these minimal energy states are unknown, it is impossible to sample Boltzmann-Gibbs measures at low temperature.

One strategy is to consider a sequence of Metropolis-Hastings transition M_n such that for any time n we have

$$\mu_{\beta_n} M_n = \mu_{\beta_n} \iff \pi_n M_n = \pi_n$$

We recall that the Markov transition M_n associated with a λ-reversible proposition transition K_n is given by

$$M_n(x,dy) = K_n(x,dy)\ a_n(x,y) + \left[1 - \int K_n(x,dz)\ a_n(x,z)\right]\ \delta_x(dy) \tag{7.15}$$

with the acceptance rate

$$a_n(x,y) = 1 \wedge e^{-\beta_n\ (V(y)-V(x))} = e^{-\beta_n\ (V(y)-V(x))_+}$$

To simplify the presentation, with start with a null inverse temperature parameter $\beta_0 = 0$, and a r.v. X_0 with distribution $\eta_0 = \mu_{\beta_0} = \lambda$. We run a series of m_1 MCMC moves with Markov transition M_1

$$X_0 \xrightarrow{\quad M_1^{m_1} \quad} X_{m_1}$$

If m_1 is sufficiently large, we expect X_{m_1} to be approximately distributed according to the invariant measure $\pi_1 = \mu_{\beta_1}$ of the transition M_1. Nevertheless, when β_1 is too large the acceptance rate $a_1(x,y) = e^{-\beta_1(V(y)-V(x))}$ is almost null for any $V(x) < V(y)$. In other words, the the sequence of M_1-MCMC moves are almost equivalent to a series of gradient-descent-type transitions. Thus, for large values of β_1 we cannot expect to have $\mathrm{Law}(X_{m_1}) \simeq \pi_1$ but for very large values of the parameter m_1.

Thus, the natural idea is to find a judicious schedule (β_n, m_n) such that the time inhomogenous model

$$X_0 \xrightarrow{M_1^{m_1}} X_{m_1} \xrightarrow{M_2^{m_2}} X_{m_1+m_2} \xrightarrow{M_3^{m_3}} X_{m_1+m_2+m_3} \longrightarrow \dots$$

explores randomly the state space with

$$\forall n \in \mathbb{N} \qquad \text{Law}(X_{m_1+\dots+m_n}) \simeq \pi_n$$

One natural idea is to introduce an intermediate acceptance-rejection mechanism every time we change the temperature parameter. For instance, initially we set

$$\widehat{X}_0 = \left\{ \begin{array}{ccc} X_0 & \text{with probability} & e^{-(\beta_1-\beta_0)V(X_0)} \\ c & \text{with probability} & 1 - e^{-(\beta_1-\beta_0)V(X_0)} \end{array} \right.$$

where c stands for some auxiliary cemetery state. Notice that for any function f on S we have

$$\begin{aligned} \mathbb{E}\left(f(\widehat{X}_0) \mid \widehat{X}_0 \neq c\right) &\propto \mathbb{E}\left(\mathbb{E}\left(f(\widehat{X}_0)\, 1_{\widehat{X}_0 \neq c} \mid X_0\right)\right) \\ &= \mathbb{E}\left(f(X_0)\, e^{-(\beta_1-\beta_0)V(X_0)}\right) \\ &\propto \int f(x)\, e^{-(\beta_1-\beta_0)V(x)}\, e^{-\beta_0 V(x)}\, \lambda(dx) \end{aligned}$$

This implies that

$$\mathbb{E}\left(f(\widehat{X}_0) \mid \widehat{X}_0 \neq c\right) \propto \int f(x)\, e^{-\beta_1 V(x)}\, \lambda(dx) \propto \pi_1(f)$$

In much the same way, recalling that $\beta_0 = 0$ we prove that

$$\mathbb{P}(\widehat{X}_0 \neq c) = \lambda\left(e^{-\beta_1 V(x)}\right) / \lambda\left(e^{-\beta_0 V(x)}\right) = \lambda\left(e^{-\beta_1 V(x)}\right)$$

If $\widehat{X}_0 = c$ then the algorithm stops. Otherwise, starting from $\widehat{X}_0 = X_0$, as before we run m_1 transitions M_1 up to some random state X_{m_1}. Notice that

$$\text{Law}(\widehat{X}_0 \mid \widehat{X}_0 \neq c) = \pi_1 \quad \Longrightarrow \quad \forall m_1 \geq 1 \quad \text{Law}(X_{m_1}) = \pi_1$$

Then, we accept-reject this state as follows

$$\widehat{X}_{m_1} = \left\{ \begin{array}{ccc} X_{m_1} & \text{with probability} & e^{-(\beta_2-\beta_1)V(X_{m_1})} \\ c & \text{with probability} & 1 - e^{-(\beta_2-\beta_1)V(X_{m_1})} \end{array} \right.$$

Arguing as above, we find that

$$\text{Law}(\widehat{X}_{m_1} \mid \widehat{X}_{m_1} \neq c,\ \widehat{X}_0 \neq c) = \pi_2$$

Similarly, we also prove that

$$\mathbb{P}(\widehat{X}_{m_1} \neq c \mid \widehat{X}_0 \neq c) = \lambda\left(e^{-\beta_2 V(x)}\right) / \lambda\left(e^{-\beta_1 V(x)}\right) = \pi_1\left(e^{-(\beta_2-\beta_1)V(x)}\right)$$

$$\Longrightarrow \quad \mathbb{P}(\widehat{X}_{m_1} \neq c) = \lambda\left(e^{-\beta_2 V(x)}\right)$$

as well as

$$\forall m_2 \geq 1 \qquad \text{Law}(X_{m_1+m_2}) = \pi_2$$

where $X_{m_1+m_2}$ stands for the random states of the model after m_2 iterations of the transition M_2, starting from $\widehat{X}_{m_1} = X_{m_1}(\neq c)$. As before, when $\widehat{X}_{m_1} = c$ the algorithm is stopped. Iterating this algorithm we obtain a sequence of perfect random samples w.r.t. the target measures π_n, as soon as the states are accepted at every acceptance-rejection transitions.

The sequential simulated annealing developed above is based on an acceptance-rejection interpretation of the product formula

$$\mu_{\beta_n}(dx) \propto \left\{ \prod_{0\leq k\leq n} h_k(x) \right\} \lambda(dx) = e^{-\beta_n V(x)} \lambda(dx) \tag{7.16}$$

with the potential functions

$$h_k(x) = e^{-(\beta_k-\beta_{k-1})V(x)} \quad \text{and the conventions} \quad \beta_0 = 0 = \beta_{-1}$$

The main drawback of this perfect sampling algorithm comes from the fact that the acceptance rate decreases exponentially fast to 0; that is, we have that

$$\mathbb{P}(\widehat{X}_{m_1+\ldots+m_n} \neq c) = \lambda\left(e^{-\beta_n V(x)}\right) \downarrow_{n\uparrow\infty} 0$$

In section 9.1.5, we design an alternative interpretation of the product representation (7.16) in terms of recycling mechanisms.

7.5.2 Sequential multilevel model

We suppose that we are given a sequence of target measures π_n defined in terms of a sequence of Boltzmann-Gibbs measures

$$\pi_n(dx) = \mu_{A_n}(dx) = \frac{1}{\mathcal{Z}_{A_n}} 1_{A_n}(x) \lambda(dx) \tag{7.17}$$

associated with sequence of non increasing subsets A_n, and some reference measure λ on some state space S. We let $K(x, dy)$ be some Markov transition such that

$$\lambda(dx)K(x, dy) \sim \lambda(dy)K(y, dx)$$

We consider the acceptance-rejection Markov transition $M_n(x, dy)$ defined by

$$\forall x \in A_n \qquad M_n(x, dy) = K(x, dy)\ a_n(x, y) + \left[1 - \int K_n(x, dz)\ a_n(x, z)\right] \delta_x(dy) \tag{7.18}$$

with the acceptance rate

$$\forall x \in A_n \qquad a_n(x, y) = 1 \wedge \left(\frac{\lambda(dy)K(y, dx)}{\lambda(dx)K(x, dy)} 1_{A_n}(y)\right) = 1 \wedge \left(\frac{\pi_n(dy)K(y, dx)}{\pi_n(dx)K(x, dy)}\right)$$

When $x \notin A_n$, we set $M_n(x, dy) = K(x, dy)$. By construction, M_n is π_n-reversible so that $\pi_n M_n = \pi_n$.

7.6 Extended Markov chain Monte Carlo models

We let $S = (S_1 \times S_2)$ be a product space, and we let π be the Boltzmann-Gibbs measure on S defined by (7.1) with a reference measure λ on S of the following form

$$\lambda(d(u_1, u_2)) = \lambda_1(du_1)\ \lambda_{1,2}(u_1, du_2)\ = \lambda_2(du_2)\ \lambda_{2,1}(u_2, du_1) \tag{7.19}$$

for some probability measures $\lambda_i \in \mathcal{P}(S_i)$ and some Markov transitions $\lambda_{i,j}$ from S_i into S_j, with $i, j \in \{1, 2\}$.

In this situation, we clearly have the disintegration formulae

$$\pi(d(u_1, u_2)) = \pi_1(du_1)\ \pi_{1,2}(u_1, du_2)\ = \pi_2(du_2)\ \pi_{2,1}(u_2, du_1) \tag{7.20}$$

with the potential functions

$$G_1(u_1) = \int \lambda_{1,2}(u_1, du_2)\ G(u_1, u_2) \quad \text{and} \quad G_2(u_2) = \int \lambda_{2,1}(u_2, du_1)\ G(u_1, u_2)$$

and the marginal measures $\pi_i = \Psi_{G_i}(\lambda_i)$ and the conditional distributions

$$\begin{aligned} \pi_{1,2}(u_1, du_2) &:= G_1(u_1)^{-1}\ \lambda_{1,2}(u_1, du_2)\ G(u_1, u_2) \\ \pi_{2,1}(u_2, du_1) &:= G_2(u_2)^{-1}\ \lambda_{2,1}(u_2, du_1)\ G(u_1, u_2) \end{aligned}$$

When we know how to sample the conditional probability measures $\pi_{1,2}$ and $\pi_{2,1}$ we can use the Gibbs-Glauber model to design an MCMC model with the invariant measure π (cf. section 7.3). In other instance, the conditional probability measures can be sample using auxiliary MCMC models.

To describe these models, we consider a Markov transition from S into itself given by

$$M((u_1, u_2), d(v_1, v_2)) = K_{u_1,1}(u_2, dv_2)\ K_{v_2,2}(u_1, dv_1)$$

with a collection of MCMC transitions $K_{u_i,i}$ from S_j into itself such that

$$\pi_{1,2}(u_1, du_2) = \int \pi_{1,2}(u_1, dv_2)\ K_{u_1,1}(v_2, du_2) \tag{7.21}$$

$$\pi_{2,1}(u_2, du_1) = \int \pi_{2,1}(u_2, dv_1)\ K_{u_2,2}(v_1, du_1) \tag{7.22}$$

Next, we check that π is an invariant measure of M.

By construction, we have

$$\begin{aligned} &\int_{u_1,u_2} \pi_1(du_1)\ \pi_{1,2}(u_1, du_2)\ K_{u_1,1}(u_2, dv_2)\ K_{v_2,2}(u_1, dv_1) \\ &= \int_{u_1} \pi_1(du_1)\ \pi_{1,2}(u_1, dv_2)\ K_{v_2,2}(u_1, dv_1) && \text{(by (7.21))} \\ &= \int_{u_1} \pi_2(dv_2)\ \pi_{2,1}(v_2, du_1)\ K_{v_2,2}(u_1, dv_1) && \text{(by (7.20))} \\ &= \pi_2(dv_2)\pi_{2,1}(v_2, dv_1) && \text{(by (7.21))} \end{aligned}$$

7.7 Continuous time embeddings

We consider a time homogeneous jump process X_t with generator

$$L(f)(x) = \lambda(x) \int [f(y) - f(x)] \, M(x, dy)$$

with some bounded intensity function $\lambda(x) \leq \lambda_{max}$ for some finite $\lambda_{max} < \infty$, and some Markov transition M. We set

$$M_\lambda(x, dy) = \frac{\lambda(x)}{\lambda_{max}} \, M(x, dy) + \left(1 - \frac{\lambda(x)}{\lambda_{max}}\right) \, \delta_x(dy) \Rightarrow L = \lambda_{max} \, [M_\lambda - Id]$$

By construction, the sg of X_t is given by

$$\begin{aligned} P_t(f)(x) &= \mathbb{E}\left(f(X_t) \mid X_0 = x\right) \\ &= \sum_{n \geq 0} \mathbb{E}\left(f(X_t) \, 1_{N_t = n} \mid X_0 = x\right) \\ &= e^{-\lambda_{max} t} \sum_{n \geq 0} \frac{(t\lambda_{max})^n}{n!} \, M_\lambda^n(f)(x) \end{aligned}$$

where N_t is a Poisson process with intensity λ_{max}.

For the unit rate function $\lambda(x) = 1$, the Markov chain X_t is the continuous time embedding of the Markov chain $\mathcal{X}_n$ with elementary transition M. These embedding techniques allow to define the continuous time version of the Metropolis-Hasting model and the Gibbs-Glauber dynamics discussed in section 7.2 and in section 7.3.

There are also several ways to transfer the stability properties of the embedded discrete generation Markov chain X_n^λ with transition M_λ to the stability of the continuous time model X_t.

Next, we present three possible routes. The first one is expressed in terms of stands for the Dobrushin ergodic coefficient $\beta(K)$ of a Markov transition K introduced in section 4.4.1. The second one is expressed in terms of the V-Dobrushin local contraction coefficient $\beta_V(M_\lambda)$ presented in definition 4.4.4. The third one is related to coupling techniques.

Whenever they exist, the invariant measures π, resp. π_λ, of the Markov transitions M, resp. M_λ, and connected for any $f \in \mathcal{B}(S)$ by the formula

$$\pi_\lambda(f) = \pi(f/\lambda) \quad \text{and} \quad \pi_\lambda L(f) = 0 \tag{7.23}$$

1. We assume that there exists some $m \geq 1$ s.t. $\beta(M_\lambda^m) < 1$. In this situation, for any $t \geq 0$, we have the exponential estimate

$$\beta(P_t) \leq \frac{1}{\beta(M_\lambda^m)^{1-1/m}} \ \exp\left[-t\lambda_{max}\left(1 - \beta(M_\lambda^m)^{\frac{1}{m}}\right)\right] \tag{7.24}$$

2. We assume that $\beta_V(M_\lambda^m) < 1$ for some $m \geq 1$ and some function $V \geq 0$. In this situation, for any $t \geq 0$, we have the exponential estimate

$$\beta_V(P_t) \leq \frac{1}{\beta_V(M_\lambda^m)^{1-1/m}} \ \exp\left[-t\lambda_{max}\left(1 - \beta_V(M_\lambda^m)^{\frac{1}{m}}\right)\right] \tag{7.25}$$

   We further assume that $0 < \lambda_{min} \leq \lambda \leq \lambda_{max}$, and M satisfies the Foster-Lyapunov condition (4.40) for some $\epsilon \in [0, 1[$, some finite $c < \infty$, and some function $W \geq 0$. In this situation, M_λ satisfies the Foster-Lyapunov condition (4.40) with

$$M_\lambda(W) \leq \left(1 - \frac{\lambda_{min}}{\lambda_{max}}\ (1 - \epsilon)\right)\ W + c \tag{7.26}$$

   In addition, if M_λ satisfies the Dobrushin local contraction condition (4.39) then there exists some function V s.t. $\beta_V(M_\lambda) < 1$.

3. We let $T_{x,y}^\lambda$ be a coupling time of two copies X_n^λ and Y_n^λ of the Markov chain with Markov transition M_λ starting at $X_n^\lambda = x$ and $Y_n^\lambda = y$. We assume that

$$\mathbb{P}\left(T_{x,y}^\lambda \geq n\right) \leq a_\lambda(x, y)\ \exp\left(-b_\lambda n\right)$$

   for some finite function $a_\lambda(x, y) < \infty$ and some positive constant $b_\lambda \in]0, 1[$. In this situation, we have

$$\begin{aligned} &\|\text{Law}(X_t \mid X_0 = x) - \text{Law}(X_t \mid X_0 = y)\|_{tv} \\ &\leq a_\lambda(x, y)\ \exp\left(-\lambda_{max} t(1 - e^{-b_\lambda})\right) \end{aligned} \tag{7.27}$$

Proof :
The r.h.s. of (7.23) is immediate, and the l.s.d. come from the following observations:

$$\begin{aligned} \pi M = \pi \Rightarrow \pi\left(\frac{1}{\lambda}(\lambda M(f) + (1 - \lambda)f)\right) &= \pi M(f) + \pi(f/\lambda) - \pi(f) \\ &= \pi(f/\lambda) = \pi_\lambda(f) \end{aligned}$$

and

$$\begin{aligned} \pi_\lambda M_\lambda = \pi_\lambda \Rightarrow \pi_\lambda(\lambda M(f)) &= \pi_\lambda(M_\lambda(f)) - \pi_\lambda((1 - \lambda)(f)) \\ &= \pi_\lambda(\lambda f) \end{aligned}$$

Now, we come to the proof of (7.24). For any function f s.t. $\text{osc}(f) \leq 1$ using theorem **??** we have

$$\text{osc}\left(M_\lambda^{nm+p}(f)\right) \leq \beta(M_\lambda^m)^n \text{osc}(M_\lambda^p(f)) \leq \beta(M_\lambda^m)^n$$

$$\begin{aligned}
\text{osc}\,(P_t(f)) &\leq e^{-\lambda_{max}t} \sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!}\ \text{osc}\,(M_\lambda^n(f)) \\
&= e^{-t\lambda_{max}} \sum_{n\geq 0} \sum_{0\leq p<m} \frac{(t\lambda_{max})^{nm+p}}{(nm+p)!}\ \text{osc}\left(M_\lambda^{nm+p}(f)\right) \\
&\leq e^{-t\lambda_{max}} \sum_{n\geq 0} \sum_{0\leq p<m} \frac{(t\lambda_{max})^{nm+p}}{(nm+p)!}\ \beta(M_\lambda^m)^n
\end{aligned}$$

When $\beta(M_\lambda^m) = 0$ the result is obvious, When $\beta(M_\lambda^m) > 0$, we observe that

$$\begin{aligned}
\frac{(t\lambda_{max})^{nm+p}}{(nm+p)!}\beta(M_\lambda^m)^n &= \frac{(t\lambda_{max}\beta(M_\lambda^m)^{\frac{1}{m}})^{nm+p}}{(nm+p)!}\ \frac{1}{\beta(M_\lambda^m)^{p/m}} \\
&\leq \frac{(t\lambda_{max}\beta(M_\lambda^m)^{\frac{1}{m}})^{nm+p}}{(nm+p)!}\ \frac{1}{\beta(M_\lambda^m)^{1-1/m}}
\end{aligned}$$

This implies that

$$\begin{aligned}
\text{osc}\,(P_t(f)) &\leq \frac{e^{-t\lambda_{max}}}{\beta(M_\lambda^m)^{1-1/m}} \sum_{n\geq 0} \frac{1}{n!}\ \left(t\lambda_{max}\beta(M_\lambda^m)^{\frac{1}{m}}\right)^n \\
&= \frac{1}{\beta(M_\lambda^m)^{1-1/m}}\ \exp\left[-t\lambda_{max}\left(1-\beta(M_\lambda^m)^{\frac{1}{m}}\right)\right]
\end{aligned}$$

This ends the proof of (7.24).

To prove (7.25) we use theorem 4.51 to check that

$$\begin{aligned}
\|P_t(x,.) - P_t(y,.)\|_V &\leq e^{-\lambda_{max}t} \sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!}\ \|M_\lambda^n(x,.) - M_\lambda^n(y,.)\|_V \\
&\leq e^{-\lambda_{max}t} \sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!}\ \beta_V(M_\lambda^n)\ \|\delta_x - \delta_x\|_V \\
&= e^{-\lambda_{max}t} \sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!}\ \beta_V(M_\lambda^n)\ (1+V(x)+V(y))
\end{aligned}$$

This implies that

$$\beta_V(P_t) \leq e^{-\lambda_{max}t} \sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!}\ \beta_V(M_\lambda^n)$$

The end of the proof of (7.25) follows the same arguments as the ones we used in the proof of (7.24). The proof of (7.26) comes from the fact that

$$\begin{aligned}
M_\lambda(W) &= \frac{\lambda}{\lambda_{max}}\ M(W) + \left(1-\frac{\lambda}{\lambda_{max}}\right)\ W \\
&\leq \left[\epsilon\ \frac{\lambda}{\lambda_{max}} + \left(1-\frac{\lambda}{\lambda_{max}}\right)\ \right] W + \frac{\lambda}{\lambda_{max}}\ c \\
&\leq \left[1-\frac{\lambda_{min}}{\lambda_{max}}(1-\epsilon)\right]\ W + c
\end{aligned}$$

The last assertion is a consequence of theorem 4.51.

The proof of (7.27) is a direct consequence of the fact that

$$\|\mathrm{Law}(X_t \mid X_0 = x) - \mathrm{Law}(X_t \mid X_0 = y)\|_{tv} = \sup_{\mathrm{osc(f)}\leq 1} |P_t(f)(x) - P_t(f)(y)|$$

and

$$\begin{aligned}
&|P_t(f)(x) - P_t(f)(y)| \\
&\leq e^{-\lambda_{max}t} \ \textstyle\sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!} \ |M_\lambda^n(f)(x) - M_\lambda^n(f)(y)| \\
&= e^{-\lambda_{max}t} \ \textstyle\sum_{n\geq 0} \frac{(t\lambda_{max})^n}{n!} \ \left\|\mathrm{Law}(X_n^\lambda \mid X_0 = x) - \mathrm{Law}(Y_n^\lambda \mid Y_0 = y)\right\|_{tv}
\end{aligned}$$

The third result is now a direct consequence of the fact that

$$\left\|\mathrm{Law}(X_n^\lambda \mid X_0 = x) - \mathrm{Law}(Y_n^\lambda \mid Y_0 = y)\right\|_{tv} \leq \mathbb{P}\left(X_n^\lambda \neq Y_n^\lambda \mid X_0 = x, Y_0 = y\right)$$

for any coupling of the chains X_n^λ and Y_n^λ. This ends the proof of the theorem. ∎

We illustrate the continuous coupling inequality (7.27) with the Markov transition $M(x, y) = 1/d$ on a finite and complete graph with d vertices $S := \{1, \ldots, d\}$. In this context, the coupling time T of two independent chains with transition M is a geometric random variable

$$\mathbb{P}(T = n) = \left(1 - \tfrac{1}{d}\right)^{n-1} \ \tfrac{1}{d}$$

This implies that

$$\mathbb{P}\left(T > n\right) = \left(1 - \frac{1}{d}\right)^n \ \sum_{p\geq n} \left(1 - \frac{1}{d}\right)^{p-n} \ \frac{1}{d} = \left(1 - \frac{1}{d}\right)^n \leq e^{-n/d} \tag{7.28}$$

Next, we assume that $\lambda_t = \lambda_{max} = \lambda$ (so that $M_\lambda = M$). Combining (7.27) with (7.28), we readily check that

$$\|\mathrm{Law}(X_t \mid X_0 = x) - \mathrm{Law}(X_t \mid X_0 = y)\|_{tv} \leq e^{-\lambda t(1-e^{-1/d})} \simeq_{d\uparrow\infty} \ e^{-t\lambda/d}$$

7.8 Gradient flow models

7.8.1 Steepest descent model

As their name indicates stochastic gradient flow models are the stochastic version of the well known steepest descent dynamical systems. Suppose we are given some manifold S. The steepest descent evolution equation in a chart $\phi \ : \ x \in S \mapsto \phi(x) \in S_\phi \subset \mathbb{R}^p$ is given by

$$\dot{\theta}_t = -(\nabla_g V)(\theta_t)$$

with the Riemannian vector field gradient defined in (6.49). For Euclidian state spaces $S = \mathbb{R}^p$ the chart reduces to the identity mapping $\phi(x) = \theta = x = \phi^{-1}(\theta)$, and $\nabla_g V = \partial V$ is the traditional gradient.

We further assume that the manifold $S = \varphi^{-1}(0)$ is the null level set of some smooth function $\varphi \ : \ x \in \mathbb{R}^{r=p+q} \mapsto \mathbb{R}^q$ s.t. $\mathrm{rank}(\partial\varphi(x)) = q$, for any $x \in \mathbb{R}^r$. In this situation, for any $1 \leq l \leq q$ we

have

$$\begin{aligned}
\frac{d}{dt}\varphi_l(\psi(\theta_t)) &= \sum_{1\leq i\leq p} \left(\partial_{\theta_i}(\varphi_l\circ\psi)\right)(\theta_t)\ \dot{\theta}_t^i \\
&= -\sum_{1\leq i,j\leq p} g^{i,j}(\theta_t) \sum_{1\leq k\leq r} \left(\partial_{x_k}\varphi_l\right)(\psi(\theta_t))\left(\partial_{\theta_i}\psi^k\right)(\theta_t)\ \left(\partial_{\theta_j}V\right)(\theta_t) \\
&= -\sum_{1\leq i,j\leq p} g^{i,j}(\theta_t)\ \underbrace{\left\langle \left(\partial\varphi_l\right)(\psi(\theta_t)), \left(\partial_{\theta_i}\psi\right)(\theta_t)\right\rangle}_{=0}\ \left(\partial_{\theta_j}V\right)(\theta_t)
\end{aligned}$$

This shows that the gradient flow keeps the state x_t in the constraint manifold at any time t as soon as we start in the desired manifold. More formally, we have

$$x_0=\psi(\theta_0)\in S \Rightarrow \forall t\geq 0 \qquad x_t=\psi(\theta_t)\in S$$

We also notice that

$$\begin{aligned}
\frac{d}{dt}V(\theta_t) &= -\sum_{1\leq i\leq p} g^{i,j}(\theta_t)\ \left(\partial_{\theta_i}V\right)(\theta_t)\ \left(\partial_{\theta_j}V\right)(\theta_t) \\
&= -\left\langle (\nabla_g V)(\theta_t), (\nabla_g V)(\theta_t)\right\rangle_{g(\theta_t)} = -\left\langle \dot{\theta}_t, \dot{\theta}_t\right\rangle_{g(\theta_t)} = -\left\|\dot{\theta}_t\right\|_{g(\theta_t)} \Rightarrow\ V(\theta_t)\downarrow
\end{aligned}$$

7.8.2 Euclidian state spaces

We start with an elementary example. The distribution

$$\pi(dx)\propto e^{-\frac{x^2}{2\sigma^2}}\ dx$$

on $\mathbb{R}$ is reversible w.r.t. the Ornstein-Uhlenbeck semigroup P_t associated with the generator

$$L(f)(x)=-\frac{x}{\sigma^2}\ \partial_x f(x)+\partial_x^2 f(x)$$

A simple way to check this claim is to rewrite the generator as follows

$$L(f)(x)=e^{\frac{x^2}{2\sigma^2}}\ \partial_x\left(e^{-\frac{x^2}{2\sigma^2}}\ \partial_x f\right)(x)$$

By a simple integration by part, we have

$$\begin{aligned}
\int e^{-\frac{x^2}{2\sigma^2}}\ f(x)\ L(g)(x)\ dx &= \int f(x)\ \partial_x\left(e^{-\frac{x^2}{2\sigma^2}}\ \partial_x g\right)(x)\ dx \\
&= -\int \partial_x f(x)\ e^{-\frac{x^2}{2\sigma^2}}\ \partial_x g(x)\ dx \\
&= \int \partial_x\left(e^{-\frac{x^2}{2\sigma^2}}\partial_x f\right)(x)\ g(x)\ dx \\
&= \int e^{-\frac{x^2}{2\sigma^2}}\ L(f)(x)\ g(x)\ dx
\end{aligned}$$

More generally, we let S be a finite set,

$$V\ :\ x=(x_i)_{i\in S}\in E=\mathbb{R}^S\mapsto V(x)\in[0,\infty[$$

a sufficiently smooth function that tends to infinity sufficiently fast when one of the coordinates of x tends to infinity, and $\alpha,\beta\in]0,\infty[$ some given parameters.

 The Boltzmann Gibbs measure

$$\pi(dx) \propto e^{-\frac{2\beta}{\sigma^2}\ V(x)}\ \lambda(dx)$$

where λ stands for the Lebesgue measure on E, is reversible w.r.t. the semigroup P_t associated with the generator

$$L = -\beta\ \nabla V \cdot \nabla + \frac{1}{2}\ \sigma^2\ \triangle \iff L(f)(x) = \frac{1}{2}\ \sigma^2 \sum_{i\in S} \partial^2_{x_i} f(x) - \beta \sum_{i\in S} \partial_{x_i} V(x)\ \partial_{x_i} f(x)$$

 Here again, a natural way to check this claim is to rewrite the generator as follows

$$L(f)(x) = \frac{1}{2}\ \sigma^2\ e^{\frac{2\beta}{\sigma^2}\ V(x)} \sum_{i\in S} \partial_{x_i} \left(e^{-\frac{2\beta}{\sigma^2}\ V(x)}\ \partial_{x_i} f \right)(x)$$

and use an integration by part to check the desired reversibility property. The stochastic gradient diffusion with generator L is given by

$$dX_t = -\beta\ \nabla V(X_t)\ dt + \sigma\ dB_t$$

where $B_t = (B^i_t)_{i\in S}$ stands for a sequence of independent Brownian motions on $\mathbb{R}$.

We consider a function a sufficiently smooth function of the form

$$W\ :\ z = (x_i, y_i)_{i\in S} \in E = (\mathbb{R}^2)^S \mapsto V(z) = U(x) + V(y) \in [0, \infty[$$

and we set

$$\begin{aligned} L_1 &:= \frac{1}{2}\ \sigma_1^2 \sum_{i\in S} \partial^2_{x_i} - \sum_{i\in S} \left[\alpha_{(1,1)} \partial_{x_i} U(x) + \alpha_{(1,2)} \partial_{y_i} V(y)\right]\ \partial_{x_i} \\ L_2 : &= \frac{1}{2}\ \sigma_2^2 \sum_{i\in S} \partial^2_{y_i} - \sum_{i\in S} \left[\alpha_{(2,1)} \partial_{x_i} U(x) + \alpha_{(2,2)} \partial_{y_i} V(y)\right]\ \partial_{y_i} \end{aligned}$$

The Boltzmann Gibbs measure

$$\pi(d(x,y)) \propto e^{-\beta_1 U(x) - \beta_2 V(y)}\ \lambda(dx)\ \lambda(dy)$$

is an invariant measure of the semigroup associated with the generator $L = L_1 + L_2$ as soon as the following conditions are met

$$\forall i = 1, 2 \qquad \alpha_{(i,i)} = \frac{1}{2}\ \sigma_i^2 \beta_i \quad \text{and} \quad \beta_1\ \alpha_{(1,2)} + \beta_2\ \alpha_{(2,1)} = 0$$

Proof:

We check this claim using the fact that

$$
\begin{aligned}
& I_1(f) \\
& := \int \; e^{-\beta_1 U(x)} \; \left[\frac{1}{2} \; \sigma_1^2 \; \partial_{x_i}^2 - \left[\alpha_{(1,1)} \partial_{x_i} U(x) + \alpha_{(1,2)} \partial_{y_i} V(y) \right] \; \partial_{x_i} \right] (f)(x) \; dx \\
& = \int \; f(x) \; \left[\frac{1}{2} \; \sigma_1^2 \; \partial_{x_i}^2 \left(e^{-\beta_1 U} \right)(x) + \alpha_{(1,1)} \; \partial_{x_i} \left(\partial_{x_i} U \; e^{-\beta_1 U} \right)(x) + \alpha_{(1,2)} \; \partial_{y_i} V(y) \; \partial_{x_i} \left(e^{-\beta_1 U} \right)(x) \right] \; dx
\end{aligned}
$$

$$
\begin{aligned}
\partial_{x_i} \left(e^{-\beta_1 U} \right)(x) &= -\beta_1 \partial_{x_i} U \; e^{-\beta_1 U} \\
\partial_{x_i} \left(\partial_{x_i} U \; e^{-\beta_1 U} \right) &= -\beta_1 \left(\partial_{x_i} U \right)^2 \; e^{-\beta_1 U} + \; \partial_{x_i}^2 U \; e^{-\beta_1 U} \\
\partial_{x_i}^2 \left(e^{-\beta_1 U} \right) &= -\beta_1 \; \partial_{x_i} \left(\partial_{x_i} U \; e^{-\beta_1 U} \right) = \beta_1^2 \left(\partial_{x_i} U \right)^2 \; e^{-\beta_1 U} - \beta_1 \; \partial_{x_i}^2 U \; e^{-\beta_1 U}
\end{aligned}
$$

This implies that

$$
\begin{aligned}
I_1(f) \; := \; & \int \; e^{-\beta_1 U(x)} \; f(x) \; \left[\frac{1}{2} \; \sigma_1^2 \; \left[\beta_1^2 \left(\partial_{x_i} U \right)^2 - \beta_1 \; \partial_{x_i}^2 U \right] \right. \\
& \left. + \alpha_{(1,1)} \; \left[-\beta_1 \left(\partial_{x_i} U \right)^2 + \; \partial_{x_i}^2 U \right] - \beta_1 \; \alpha_{(1,2)} \; \partial_{y_i} V(y) \partial_{x_i} U \; \right] \; dx \\
= \; & \int \; e^{-\beta_1 U(x)} \; f(x) \; \left[\left(\beta_1 \; \left(\partial_{x_i} U \right)^2 - \partial_{x_i}^2 U \right) \left(\frac{1}{2} \; \sigma_1^2 \beta_1 - \alpha_{(1,1)} \right) - \beta_1 \; \alpha_{(1,2)} \; \partial_{y_i} V(y) \partial_{x_i} U(x) \right]
\end{aligned}
$$

By symmetry arguments, we also have

$$
\begin{aligned}
& I_2(f) \\
& := \int \; e^{-\beta_2 V(y)} \; \left[\frac{1}{2} \; \sigma_2^2 \; \partial_{y_i}^2 - \left[\alpha_{(2,2)} \partial_{y_i} V(y) + \alpha_{(2,1)} \partial_{x_i} U(x) \right] \; \partial_{y_i} \right] (f)(y) \; dy \\
& = \int \; e^{-\beta_2 V(y)} \; f(y) \; \left[\left(\beta_2 \; \left(\partial_{y_i} V \right)^2 - \partial_{y_i}^2 V \right) \left(\frac{1}{2} \; \sigma_2^2 \beta_2 - \alpha_{(2,2)} \right) - \beta_2 \; \alpha_{(2,1)} \; \partial_{x_i} U(x) \partial_{y_i} V(y) \right]
\end{aligned}
$$

The end of the proof is now clear. This completes the proof of the proposition. ∎

7.8.3 Langevin diffusions on manifolds

We consider the projected diffusion model (6.30) associated with a gradient $b = \partial V$ of some smooth function $V \; : \; \mathbb{R}^r \mapsto \mathbb{R}$.

When $\sigma = Id$, we have $\mathbb{H}_\sigma = \mathbb{H}$ and $\pi(x)\partial V(x) = \nabla V(x)$ so that (6.30) can be interpreted as the projection on the manifold of the Langevin diffusion; that is, we have that

$$\begin{aligned} dX_t &= \pi(X_t)\ (-\partial V(X_t)dt + dB_t) - \frac{1}{2}\ \mathbb{H}(X_t)\ dt \\ &= -\nabla V(X_t)\ dt + \left[\pi(X_t)\ dB_t - \frac{1}{2}\ \mathbb{H}(X_t)\ dt\right] \end{aligned} \tag{7.29}$$

In this particular situation, the generator (6.31) of X_t is given by

$$\begin{aligned} L(F) &= \frac{1}{2}\ \Delta F - \langle \nabla V, \nabla F\rangle \quad \left(\quad \text{with} \quad \Delta F = \text{tr}\left(\nabla^2 F\right)\right) \\ &= \frac{1}{2}\ e^{2V}\ \text{div}\left(e^{-2V}\ \nabla F\right) \end{aligned}$$

The divergence formulation given above is checked using the fact that

$$\nabla\left[e^{-2V}\ \nabla F\right] = \nabla\left[e^{-2V}\right]\ \left[\nabla F\right]^T + e^{-2V}\ \nabla\left[\nabla F\right]$$

and

$$\nabla\left[e^{-2V}\right] = -2\ e^{-2V}\ \nabla V$$

so that

$$\begin{aligned} \text{div}\left(e^{-2V}\ \nabla F\right) &= \text{tr}\left(\nabla\left[e^{-2V}\ \nabla F\right]\right) \\ &= -2\ e^{-2V}\ \text{tr}\left(\nabla V\ \left[\nabla F\right]^T\right) + e^{-2V}\ \text{tr}\left(\ \nabla\left[\nabla F\right]\right) \\ &= -2\ e^{-2V}\ \langle \nabla V, \nabla F\rangle + e^{-2V}\ \Delta F \end{aligned}$$

The end of the proof is now clear.

For any smooth function F with compact support and any smooth vector field $W \in T(S)$, we have

$$\int_S F\ \text{div}\,(W)\ d\mu_S = -\int_S \langle W, \nabla F\rangle\ d\mu_S$$

where μ_S stands for the volume measure on S. This integration by part divergence theorem (6.83) is proved in section 6.6.3. We let η be the Bolzmann-Gibbs measure on S defined by

$$d\eta = \frac{1}{\mathcal{Z}}\ e^{-2V}\ d\mu_S$$

where $\mathcal{Z}$ stands for some normalizing constant. Here we have implicitly assumed that $e^{-2V}\ d\mu_S \in]0, \infty[$. In this notation, we have

$$\begin{aligned} 2\ \mathcal{Z}\ \int_S F_1\ L(F_2)\ d\eta &= \int_S F_1\ e^{2V}\ \text{div}\left(e^{-2V}\ \nabla F_2\right)\ e^{-2V}\ d\mu_S \\ &= \int_S F_1\ \text{div}\left(e^{-2V}\ \nabla F_2\right)\ d\mu_S \\ &= -\int_S \left\langle e^{-2V}\ \nabla F_2, \nabla F_1\right\rangle\ d\mu_S = -\int_S \langle \nabla F_1, \nabla F_2\rangle\ e^{-2V}\ d\mu_S \end{aligned}$$

This implies the reversibility property of L w.r.t. η. More precisely, we have the following formula

$$\int_S F_1\ L(F_2)\ d\eta = -\frac{1}{2}\ \int_S \langle \nabla F_1, \nabla F_2\rangle\ d\eta = \int_S L(F_1)\ F_2\ d\eta$$

7.8.4 Riemannian Langevin diffusions

We consider the projected Langevin diffusion model (7.29), and we set $\phi(X_t) = \Theta_t$. In this situation, we have

$$dX_t^k = -(\nabla\, \chi^k)^T(X_t)\partial V(X_t)\; dt + \frac{1}{2}\; \Delta(\chi^k)(X_t)\; dt + (\nabla\, \chi^k)^T(X_t)\; dB_t$$

Arguing as in the proof of (6.80), using Ito formula we have

$$\begin{aligned} d\phi^i(X_t) &= -\sum_{1\leq k\leq r} \left(\partial_{x_k}\phi^i\right)(X_t)(\nabla\, \chi^k)^T(X_t)\; \partial V(X_t)\; dt + \frac{1}{2}\; \left(\Delta\phi^i\right)(X_t)dt + \left(\nabla\phi^i\right)^T(X_t)dB_t \\ &= -\left(\nabla\phi^i\right)^T(X_t)(\partial V)(X_t)\; dt + \frac{1}{2}\; \left(\Delta\phi^i\right)(X_t)dt + \left(\nabla\phi^i\right)^T(X_t)\; dB_t \quad (\Leftarrow (6.79)) \end{aligned}$$

Using the fact that $X_t = \psi(\Theta_t)$, we arrive at the equation

$$\forall 1 \leq i \leq q \quad d\Theta_t^i = \left[-\left(\nabla\phi^i\right)_\psi^T(\Theta_t)\; (\partial V)_\psi(\Theta_t) + \frac{1}{2}\; \left(\Delta\phi^i\right)_\psi(\Theta_t)\right]\; dt + \left(\nabla\phi^i\right)_\psi^T(\Theta_t)\; dB_t$$

Using (6.65) and (6.66) the expression of these Riemannian Langevin equation in terms of the Riemannian inner product g is given by

$$\begin{aligned} \left(\Delta\phi^i\right)_\psi &= \sum_{1\leq j\leq p} \frac{1}{\sqrt{\det(g)}}\; \partial_{\theta_j}\left(\sqrt{\det(g)}\; g^{i,j}\right) \\ \left(\nabla\phi^i\right)_\psi^T\; (\partial V)_\psi &= \sum_{1\leq j\leq p} g^{i,j}\; \left\langle \left(\partial_{\theta_j}\psi\right), (\partial V)_\psi \right\rangle \\ d\Theta_t^i d\Theta_t^j &= \left(\nabla\phi^i\right)_\psi^T(\Theta_t)\; dB_t dB_t^T\; \left(\nabla\phi^j\right)_\psi(\Theta_t) = \left\langle \left(\nabla\phi^i\right)_\psi(\Theta_t), \left(\nabla\phi^j\right)_\psi(\Theta_t)\right\rangle dt \\ &= g^{i,j}(\Theta_t)\; dt = \left\langle \sum_{1\leq k\leq p} \sqrt{g^{-1}}^i{}_k(\Theta_t)\; dB_t^k, \sum_{1\leq l\leq p} \sqrt{g^{-1}}^j{}_l(\Theta_t)\; dB_t^l \right\rangle \end{aligned}$$

with

$$\begin{aligned} \left\langle \left(\partial_{\theta_j}\psi\right), (\partial V)_\psi(\theta)\right\rangle &= \sum_{1\leq k\leq r} \left(\partial_{\theta_j}\psi^k\right)(\theta)\; (\partial_{x_k} V)_\psi(\theta) \\ &= \partial_{\theta_j}\left(V\circ\psi\right)(\theta) = \left(\partial_{\theta_j} U\right)(\theta) \quad \text{with} \quad U = V\circ\psi \end{aligned}$$

This yields

$$\begin{aligned} d\Theta_t^i &= -\sum_{1\leq j\leq p} g^{i,j}\; \left(\partial_{\theta_j} U\right)(\Theta_t)\; dt \\ &\quad + \left[\sum_{1\leq k\leq p} \sqrt{g^{-1}}^i{}_k(\Theta_t)\; dB_t^k + \frac{1}{2}\; \sum_{1\leq j\leq p}\; \frac{1}{\sqrt{\det(g(\Theta_t))}}\; \partial_{\theta_j}\left(\sqrt{\det(g)}\; g^{i,j}\right)(\Theta_t)\; dt\right] \\ &= -\left(\nabla_g U\right)^i(\Theta_t)\; dt + d\overline{B}_t^i \end{aligned}$$

with the d-dimensional Brownian motion $\overline{B}_t$ on the Riemannian manifold defined for any $1 \leq i \leq p$ by

$$d\overline{B}_t^i = \sum_{1\leq k\leq p} \sqrt{g^{-1}}^i{}_k(\Theta_t)\; dB_t^k + \frac{1}{2}\; \sum_{1\leq j\leq p}\; \frac{1}{\sqrt{\det(g(\Theta_t))}}\; \partial_{\theta_j}\left(\sqrt{\det(g)}\; g^{i,j}\right)(\Theta_t)\; dt$$

Notice that

$$\frac{1}{2}\ \sum_{1\leq j\leq p}\ \frac{1}{\sqrt{\det(g)}}\ \partial_{\theta_j}\left(\sqrt{\det(g)}\ g^{i,j}\right)$$

$$=\frac{1}{2}\ \sum_{1\leq j\leq p}\ \partial_{\theta_j}\left(g^{i,j}\right)+\frac{1}{4}\ \sum_{1\leq j\leq p}\ g^{i,j}\ \mathrm{tr}\left(g^{-1}\partial_{\theta_j}g\right)$$

This shows that

$$d\Theta_t=-\left(\nabla_g U\right)(\Theta_t)\ dt+d\overline{B}_t \tag{7.30}$$

Alternatively, in terms of the potential function U' defined by

$$e^{-2U'}:=e^{-2U}\ \sqrt{\det(g)}\Longleftrightarrow U'=U-\frac{1}{4}\log\det(g(\theta))$$

we have

$$d\Theta_t=\overbrace{\left[-\left(\nabla_g U\right)+\frac{1}{4}\sum_{1\leq j\leq p}\ g^{i,j}\ \mathrm{tr}\left(g^{-1}\partial_{\theta_j}g\right)\right]}^{=-(\nabla_g U')}(\Theta_t)dt \tag{7.31}$$

$$+\frac{1}{2}\ \sum_{1\leq j\leq p}\ \partial_{\theta_j}\left(g^{i,j}\right)(\Theta_t)\ dt+\sum_{1\leq k\leq p}\ \sqrt{g^{-1}}^{i}_{k}(\Theta_t)\ dB^k_t$$

Arguing as in (6.77), and recalling that

$$\langle\partial f,\nabla_g U\rangle=\langle\nabla_g f,\nabla_g U\rangle_g$$

the infinitesimal generator $\mathcal{L}$ of Θ_t is given by

$$\begin{aligned}\mathcal{L}(f)&=-\langle\nabla_g f,\nabla_g U\rangle_g+\frac{1}{2}\ \Delta_g(f)\\&=-\langle\nabla_g f,\nabla_g U\rangle_g+\frac{1}{2}\ \mathrm{div}_g\left(\nabla_g(f)\right)\\&=-\langle\nabla_g f,\nabla_g U\rangle_g+\frac{1}{2}\sum_{1\leq i\leq p}\frac{1}{\sqrt{\det(g)}}\ \partial_{\theta_i}\left(\sqrt{\det(g)}\ (\nabla_g f)^i\right)\quad(\Leftarrow(6.72))\\&=\frac{1}{2}\ e^{2U}\sum_{1\leq i\leq p}\frac{1}{\sqrt{\det(g)}}\ \partial_{\theta_i}\left(e^{-2U}\sqrt{\det(g)}\ (\nabla_g f)^i\right)\end{aligned}$$

This formula can be rewritten in the more synthetic form

$$\mathcal{L}(f)=\frac{1}{2}\ e^{2U}\ \ \mathrm{div}_g\left(e^{-2U}\ \nabla_g(f)\right)$$

We consider the Riemannian volume measure μ_g and the Boltzmann-Gibbs measure η on S_ψ defined by

$$\mu_g(d\theta)=\sqrt{\det(g(\theta))}\ d\theta\quad\text{and}\quad\eta(d\theta)=\frac{1}{\mathcal{Z}}\ e^{-2U(\theta)}\ \mu_g(d\theta)=\frac{1}{\mathcal{Z}}\ e^{-2U'(\theta)}\ d\theta$$

with the normalizing constant

$$\mathcal{Z}=\int\ e^{-2U(\theta)}\ \mu_g(d\theta)=\int\ e^{-2U'(\theta)}\ d\theta$$

For any smooth functions f_1, f_2 with compact support, using a simple integration by part we have

$$\begin{aligned}\int f_1(\theta)\ \mathcal{L}(f_2)(\theta)\ e^{-2U(\theta)}\ \mu_g(d\theta) &= -\frac{1}{2}\sum_{1\leq i\leq p}\int \partial_{\theta_i}(f_1)(\theta)\ (\nabla_g f_2)^i(\theta)\ e^{-2U(\theta)}\ \mu_g(d\theta)\\ &= -\frac{1}{2}\sum_{1\leq i\leq p}\int \langle(\nabla_g f_1)(\theta),(\nabla_g f_2)(\theta)\rangle_{g(\theta)}\, e^{-2U(\theta)}\ \mu_g(d\theta)\end{aligned}$$

This shows that $\mathcal{L}$ is reversible w.r.t. η; that is we have that

$$\begin{aligned}\int f_1(\theta)\ \mathcal{L}(f_2)(\theta)\ \eta(d\theta) &= -\frac{1}{2}\sum_{1\leq i\leq p}\int \langle(\nabla_g f_1)(\theta),(\nabla_g f_2)(\theta)\rangle_{g(\theta)}\,\eta(d\theta)\\ &= \int \mathcal{L}(f_1)(\theta)\ f_2(\theta)\ \eta(d\theta)\end{aligned}$$

7.9 Metropolis-adjusted Langevin models

The choice of the time discretization schemes is extremely important. For instance, a simple Euler type discretization model may fail to transfer the desired regularity properties of the continuous model to the discrete time one.

We illustrate this assertion with a discussion on an overdamped Langevin diffusion, on an energy landscape associated with a given energy function $V \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}_+)$ on $E = \mathbb{R}^d$, for some $d \geq 1$. This model is defined by the following diffusion equation

$$dX_t = -\beta\ \nabla V(X_t) + \sqrt{2}\ dW_t \tag{7.32}$$

where ∇V denotes the gradient of V, β an inverse temperature parameter, and W_t a standard Brownian motion on $\mathbb{R}^d$. The infinitesimal generator associated with this continuous time process is given by the second order differential operator

$$L_\beta = -\beta\ \nabla V \cdot \nabla + \triangle$$

Under some regularity conditions on V, the diffusion X'_t is geometrically ergodic with an invariant measure given by

$$d\pi_\beta = \frac{1}{\mathcal{Z}_\beta}\ e^{-\beta V}\ d\lambda$$

where λ stands for the Lebesgue measure on $\mathbb{R}^d$, and $\mathcal{Z}_\beta$ is a normalizing constant.

As usual, in the continuous time framework, to get some feasible solution we need to introduce a time discretization scheme. To this end, we let $\mathcal{W}_{n+1}$ be a sequence of centered and reduced Gaussian variables on $\mathbb{R}^d$.

Firstly, starting from some random state $\mathcal{X}_n$, we propose a random state $\mathcal{Y}_{n+1}$ using the Euler scheme

$$\mathcal{Y}_{n+1} = \mathcal{X}_n - \beta\ \nabla V(\mathcal{X}_n)/m + \sqrt{2/m}\ \mathcal{W}_{n+1} \tag{7.33}$$

Then, we accept this state $\mathcal{X}_{n+1} = \mathcal{Y}_{n+1}$ with probability

$$1 \wedge \left(e^{-\beta(V(\mathcal{Y}_{n+1})-V(\mathcal{X}_n))} \times \frac{p_m(\mathcal{Y}_{n+1},\mathcal{X}_n)}{p_m(\mathcal{X}_n,\mathcal{Y}_{n+1})}\right)$$

Otherwise, we stay in the same location $\mathcal{X}_{n+1} = \mathcal{X}_n$.

In the above display, the function p_m stands for the probability density of the Euler scheme proposition

$$p_m(x,y) = \frac{1}{(4\pi/m)^{d/2}} \ \exp\Big(-\frac{m}{4}\,\|y - x + \beta\ \nabla V(x)/m\|^2\Big)$$

The resulting Markov chain model X_n is often referred to as the Metropolis-adjusted Langevin algorithm (*abbreviated (MALA)*). One of the main advantages of the above construction is that the Markov chain X_n is reversible w.r.t. to π_β, and it has the same fixed point π_β as the continuous time model. Without the acceptance-rejection rate, the Markov chain (7.33) reduces to the standard Euler approximation of the Langevin diffusion model (7.32). In this situation, the Markov chain may even fail to be ergodic, when the vector field ∇V is not globally Lipschitz [512, 438]. We refer the reader to [72, 295, 512, 513, 590] for further details on the stochastic analysis of these Langevin diffusion models. We also mention that the Euler scheme diverges in many situations, even for uniformly convex functions V. At the cost of some additional computational effort, a better idea is to replace (7.33) by the implicit backward Euler scheme given by

$$\mathcal{Y}_{n+1} + \beta\ \nabla V(\mathcal{Y}_{n+1})/m = \mathcal{X}_n + \sqrt{2/m}\ \mathcal{W}_{n+1}$$

Chapter 8

Some illustrations

8.1 Bayesian inference

8.1.1 Disintegration formulae and Gibbs sampling

 We consider a couple of random variables (Θ, X) on some state space $S = (\Xi \times E)$ with some distribution of the form

$$\pi(d(\theta, x)) = \pi_1(d\theta)\ L_{1,2}(\theta, dx) = \pi_2(dx)\ L_{2,1}(x, d\theta) \tag{8.1}$$

The first marginal measure π_1 is called the prior distribution, and the distribution $L_{2,1}(p, d\theta)$ is called the posterior distribution of Θ given the observation $X = x$. Note that $L_{1,2}(\theta, dx)$ coincides with the conditional distribution of the r.v. X given $\Theta = \theta$.

 When $L_{1,2}(\theta, dx)$ has some density $x \mapsto l_x(\theta) \propto G_x(\theta)$ w.r.t. some reference measure $\lambda(dx)$ on E, the posterior distribution takes the form

$$L_{2,1}(x, d\theta) = \Psi_{G_x}(\pi_1)(d\theta) \tag{8.2}$$

for λ-almost every $x \in E$. The function G_x is called the likelihood function of the observation variable $X = x$ given the value of the parameter $\Theta = \theta$. In these settings, the Boltzmann-Gibbs transformation is also called the Bayes' rule.

In Bayesian literature the disintegration formula (8.1) and the Boltzmann-Gibbs transformation (8.2) are often written with some abusive notation in the following form

$$p(\theta, x) = p(x|\theta)\ p(\theta)\ \ =\ \ p(\theta|x)\ p(x)$$

and

$$p(\theta|x) = \frac{1}{p(x)}\ p(\theta|x)\ p(\theta) \quad \text{with the normalizing constant} \quad p(x) := \int p(x|\theta)\ p(\theta)\ d\theta$$

In this notation, we have

$$\pi_1(d\theta) = p(\theta)d\theta \quad \pi_2(dx) = p(x)dx \quad L_{1,2}(\theta, dx) = p(x|\theta)dx \quad \text{and} \quad L_{2,1}(x, d\theta) = p(\theta|x)d\theta$$

with some hypothetic reference measures $d\theta$ and dx, and of course for different functions $p(.)$ and $p(.\mid .)$ depending on the arguments x and θ used in the formulae.

 The Gibbs sampler associated with the disintegration formulae (8.1) is defined by a Markov chain $\mathcal{X}_k = (\Theta_k, X_k) \in S = (\Theta \times E)$ with transitions

$$\mathbb{P}\left((\Theta_k, X_k) \in d(\theta_k, x_k) \mid (\Theta_{k-1}, X_{k-1})\right) = L_{2,1}(X_{k-1}, d\theta_k)\ L_{1,2}(\theta_k, dX_k) \tag{8.3}$$

We illustrate these Bayesian models with an elementary example. We let π_1 be the Beta distribution with parameter $(a, b) \in [1, \infty[^2$ on $\Xi := [0, 1]$, and $L_{1,2}$ be the Binomial type Markov transition with parameters (n, θ) from $\Xi = [0, 1]$ into $\{0, \ldots, n\} \subset E = \mathbb{R}$; more formally, we have

$$\begin{array}{rcl@{\quad}rcl} \pi_1(d\theta) & \propto & \theta^{a-1}\ (1-\theta)^{b-1}\ 1_{[0,1]}(\theta)\ d\theta & \text{and} & L_{1,2}(\theta, dx) & = & \sum_{0\le p\le n} \left(\begin{array}{c} n \\ p \end{array}\right) \theta^p\ (1-\theta)^{n-p}\ \delta_p(dx) \\ & \propto & \theta^{a-1}\ (1-\theta)^{b-1}\ \mu(d\theta) & & & = & \theta^x\ (1-\theta)^{n-x}\ \lambda(dx) \end{array} \tag{8.4}$$

with the reference measures

$$\mu(d\theta) = 1_{[0,1]}(\theta)\ d\theta \quad \text{and} \quad \lambda(dx) = \sum_{0\le k\le n} \left(\begin{array}{c} n \\ k \end{array}\right) \delta_k(dx)$$

By construction, we have that

$$\begin{array}{rcl} \pi(d(\theta, x)) & \propto & \theta^{a+x-1}\ (1-\theta)^{b+(n-x)-1}\ \mu(d\theta)\ \lambda(dx) \\ & \propto & \theta^{a+x-1}\ (1-\theta)^{b+(n-x)-1}\ \mu(d\theta) \times (\pi_1 L_{1,2})(dx) \end{array}$$

This implies that $\pi_1 L_{1,2} = \pi_2$ and $L_{2,1}$ is the Beta type Markov transition given by

$$\forall 0 \le p \le n \qquad L_{2,1}(p, d\theta) \propto \theta^{a+p-1}\ (1-\theta)^{b+(n-p)-1}\ \mu(d\theta)$$

In other words, in terms of Bayes' rule we have

$$L_{2,1}(x, d\theta) = \Psi_{G_x}(\pi_1)(d\theta) \quad \text{with the likelihood function} \quad G_x(\theta) := \theta^x\ (1-\theta)^{n-x}$$

The prior Beta distribution with parameter (a, b) is sometimes written $\pi_1 = \mathit{Beta}(a, b)$, and the Binomial conditional observation distribution $L_{1,2}(\theta, .) = \mathit{Binomial}(n, \theta)$. The Bayesian learning model is often written in the more synthetic form

$$\left\{\begin{array}{rcl} \Theta & \sim & \mathit{Beta}(a, b) \\ X \mid \Theta & \sim & \mathit{Binomial}(n, \Theta) \end{array}\right. \implies \Theta \mid X \ \sim\ \mathit{Beta}(a + X, b + (n - X)) \tag{8.5}$$

In some instances, the desired prior distributions $\text{Law}(\Theta \mid X)$ have an explicit form and the desired estimator of Θ given the observation X, such as the $\mathbb{E}(\Theta \mid X)$ for one dimensional problems, can be computed directly. In more general situations, another level of approximation is needed. One natural strategy is to use the MCMC methodologies.

For instance, in the elementary example discussed above, the Gibbs transition (8.3) reduces to the sampling of a Beta r.v. Θ_k with paramaters $((a + X_{k-1}), b + (n - X_{k-1}))$ and a Binomial r.v. X_k with parameters (n, Θ_k). It is important to notice that it is not essential to determine the marginal distribution $\pi_2 = \pi_1 L_{1,2}$ of the random variable X to define and to run the Gibbs sampler.

Next, we discuss the parameter inference problem associated with this Bayesian model. For $n = 1$, it is readily checked that $L_{1,2}$ is the Bernoulli type Markov transition from $\Xi \in [0, 1]$ into $\{0, 1\}$ given by

$$L^{(1)}_{1,2}(\theta, dx) = (1-\theta)\ \delta_0(dx) + \theta\ \delta_1(dx) \tag{8.6}$$

and the corresponding posterior distribution $L_{2,1}^{(1)}$ is the Beta type Markov transition given by

$$\forall \epsilon \in \{0,1\} \qquad L_{2,1}^{(1)}(\epsilon, d\theta) = \Psi_{g_\epsilon}(\pi_1)(d\theta) \quad \text{with} \quad g_\epsilon(\theta) := \theta^\epsilon \ (1-\theta)^{1-\epsilon}$$

The learning process associated with a sequence of observations $\epsilon := (\epsilon_k)_{1 \leq k \leq n} \in \{0,1\}^n$ is represented by a successive application of the Bayes' rule

$$\left(\Psi_{g_{\epsilon_n}} \circ \ldots \Psi_{g_{\epsilon_1}}\right)(\pi_1) = \Psi_{G_{p_n(\epsilon)}}(\pi_1)$$

with the potential function $G_{p_n(\epsilon)}$ given by

$$p_n(\epsilon) = \sum_{1 \leq k \leq n} \epsilon_k \implies G_{p_n(\epsilon)}(\theta) = \prod_{1 \leq k \leq n} g_{\epsilon_k}(\theta) = \theta^{p_n(\epsilon)} \ (1-\theta)^{n - p_n(\epsilon)}$$

Given $\Theta = \theta$, we let $B := (B_k)_{1 \leq k \leq n}$ be a sequence of conditionally independent Bernoulli r.v. with common distribution (8.6). We have shown that

$$\mathbb{P}\left(\Theta \in d\theta \mid B = \epsilon\right) = \Psi_{G_{p_n(\epsilon)}}(\pi_1)(d\theta) = L_{2,1}(p_n(\epsilon), d\theta) = \mathbb{P}\left(\Theta \in d\theta \mid X = p_n(\epsilon)\right)$$

8.1.2 Conjugate priors, likelihood and posteriors

The examples discussed above show that the prior distribution $\pi_1(d\theta)$ and the posterior distribution $L_{2,1}(x, d\theta) = \Psi_{G_x}(\pi_1)(d\theta)$, resp. $L_{2,1}^{(1)}(\epsilon, d\theta) = \Psi_{g_\epsilon}(\pi_1)(d\theta)$, are in the same class of Beta distributions. In Bayesian statistics, this property is interpreted as a conjugacy relation between distributions w.r.t. some class of likelihood functions.

We often say that a class of prior distributions $\mathcal{P}$ is conjugate to a class of likelihood functions $\mathcal{G}$ if we have

$$\forall G \in \mathcal{G} \qquad \Psi_G(\mathcal{P}) \subset \mathcal{P} \tag{8.7}$$

To be more precise, we equip the set E with a reference measure λ, and we suppose we are given a set of prior distributions of the following form

$$\mathcal{P} \quad = \quad \{ \ \nu_h \in \mathcal{P}(\Xi) \ : \ h \in \mathcal{H} \}$$

where $\mathcal{H}$ stands for a set of indexes, and a class of likelihood functions

$$\mathcal{G} = \Big\{ \ G \ : \ (x,\theta) \mapsto G_x(\theta) \ : \ \forall \theta \in \Xi \qquad \int \ \lambda(dx) \ G_x(\theta) = 1 \ \text{ and}$$

$$\forall h \in \mathcal{H} \qquad L^{(h)}(x, d\theta) := \Psi_{G_x}(\nu_h)(d\theta) \ \text{ is Markov operator from } E \text{ into } \Xi \ \Big\} \tag{8.8}$$

The state $\mathcal{H}$ is often called the set of (prior) hyper-parameters to avoid confusions with the parameters we want to make inference about.

We say that $\mathcal{P}$ is conjugate to the class of likelihood functions $\mathcal{G}$ if there exists some conjugacy mapping

$$H \ : \ (h,x) \in (\mathcal{H} \times E) \mapsto H_x(h) \in \mathcal{H} \quad \text{s.t.} \quad \forall h \in \mathcal{H} \quad \forall G \in \mathcal{G} \qquad \Psi_{G_x}(\nu_h) = \nu_{H_x(h)} \tag{8.9}$$

In this notation, the disintegration formula (8.1) takes the form

$$\pi_h(d(\theta,x)) := \underbrace{\nu_h(d\theta)}_{=\pi_1(d\theta)}\ \underbrace{G_x(\theta)\ \lambda(dx)}_{L_{1,2}(\theta,dx)} = \underbrace{\nu_h(G_x)\ \lambda(dx)}_{=\pi_2(dx)}\ \underbrace{\Psi_{G_x}(\nu_h)(d\theta)}_{=L_{2,1}(x,d\theta)}$$

For instance, the example discussed above show that the Beta distribution is conjugate to itself w.r.t. the Binomial likelihood function G_x, resp. w.r.t. the Bernoulli function g_ϵ.

$$\forall h=(a,b)\in\mathcal{H}=[1,\infty[^2 \qquad \nu_h(d\theta)\propto\theta^{a-1}\ (1-\theta)^{b-1}\ 1_{[0,1]}(\theta)\ d\theta\in\mathcal{P}(\Xi)=\mathcal{P}([0,1])$$

and

$$\forall x\in\{0,\ldots,n\}\subset E=\mathbb{R} \qquad G_x(\theta)=\theta^x\ (1-\theta)^{n-x}$$

In this situation, the conjugacy mapping is given by

$$H_x(a,b)=(a+x,b+(n-x))=(a,b)+(x,n-x)$$

and (8.5) takes the following form

$$\left\{\begin{array}{rcl}\Theta & \sim & Beta(h)\\ X\mid\Theta & \sim & Binomial(n,\Theta)\end{array}\right. \implies \Theta\mid X\ \sim\ Beta(H_X(h)) \tag{8.10}$$

The class of Gaussian prior distributions is also conjugate to itself w.r.t. Gaussian likelihood functions *with known covariance matrices.*

To be more precise, we denote by $\mathcal{N}(m,R)$ the Gaussian distribution on a d-dimensional space $\mathbb{R}^d$ with mean column vector $m\in\mathbb{R}^d$ and covariance matrix $R\in\mathbb{R}^{d\times d}$

$$\mathcal{N}(m,R)(dx)=g_{(m,R)}(x)\ dx \tag{8.11}$$

with

$$g_{(m,R)}(x):=\frac{1}{(2\pi)^{d/2}\sqrt{|R|}}\ \exp\left[-2^{-1}(x-m)'R^{-1}(x-m)\right]$$

where $|R|$ stands for the determinant of R.

We assume that R is in the set Σ_d of invertible covariance matrices. In the above display, dx stands for the Lebesgue measure on $\mathbb{R}^d$. In this notation, we have

$$\Psi_{g_{(m_0,R_0)}}\left(\mathcal{N}(m_1,R_1)\right)=\mathcal{N}(m_2,R_2)$$

with

$$\begin{array}{rcl} m_2 & = & R_1(R_0+R_1)^{-1}m_0+R_0(R_0+R_1)^{-1}m_1\\ & = & (R_0^{-1}+R_1^{-1})^{-1}R_0^{-1}x+(R_0^{-1}+R_1^{-1})^{-1}R_1^{-1}m_1 \quad \text{and} \quad R_2^{-1}=R_0^{-1}+R_1^{-1}\end{array}$$

We check this claim using the fact that

$$(x-m_0)'R_0^{-1}(x-m_0)+(x-m_1)'R_1^{-1}(x-m_1)$$

$$=x'R_2^{-1}x-2x'R_2^{-1}\left(R_2R_0^{-1}m_0+R_2R_1^{-1}m_1\right)+\text{some function of }(m_i,R_i)_{i=0,1}$$

$$=(x-m_2)'\,R_2^{-1}\,(x-m_2)+\text{some function of }(m_i,R_i)_{i=0,1}$$

The last assertion comes from the fact that

$$(R_0^{-1}+R_1^{-1})^{-1}R_0^{-1} = (R_0(R_0^{-1}+R_1^{-1}))^{-1} = (I+R_0R_1^{-1})^{-1} = ((R_0+R_1)R_1^{-1})^{-1} = R_1\ (R_0+R_1)^{-1}$$

and by symmetry arguments we conclude that

$$R_2R_0^{-1}m_0 + R_2R_1^{-1}m_1 = R_1(R_0+R_1)^{-1}m_0 + R_0(R_0+R_1)^{-1}m_1$$

For any fixed precision matrix $P_0 = R_0^{-1} \in \Sigma_0$ the set of prior distributions is given by

$$\forall h = (m_1, R_1^{-1}) := (m_1, P_1) \in \mathcal{H} = (\mathbb{R}^d \times \Sigma_d) \qquad \nu_h := \mathcal{N}(m_1, R_1) \in \mathcal{P}(\Xi) = \mathcal{P}(\mathbb{R}^d)$$

and the likelihood functions are defined by the Gaussian densities

$$\forall x \in E = \mathbb{R}^d \qquad G_x(\theta) = g_{(x,R_0)}(\theta) \tag{8.12}$$

In this notation, if we set $N(m, R^{-1}) = \mathcal{N}(m, R)$, then we have

$$\Psi_{G_x}(\nu_h) = \nu_{H_x(h)} \Longleftrightarrow \Psi_{G_x}(N(m_1, P_1)) = N(H_x(m_1, P_1))$$

with the conjugacy mapping (8.9) is given by

$$H_x(m_1, P_1) \quad = \quad \left((P_0+P_1)^{-1}P_0\ x + (P_0+P_1)^{-1}P_1\ m_1, P_0+P_1\right)$$

The corresponding Bayesian learning model (8.5) is given by

$$\left\{\begin{array}{rcl} \Theta & \sim & N(m_1, P_1) \\ X \mid \Theta & \sim & N(\Theta, P_0) \end{array}\right. \quad \Longrightarrow \quad \Theta \mid X \ \sim \ N(H_X(m_1, P_1)) \tag{8.13}$$

It is now readily check that

$$H_{x_2}(H_{x_1}(m_1, P_1)) \quad = \quad \left((2P_0+P_1)^{-1}P_0\ (x_1+x_2) + (2P_0+P_1)^{-1}P_1\ m_1,\ 2P_0+P_1\right)$$

Iterating the argument, we find that

$$\left(\Psi_{G_{x_n}} \circ \ldots \circ \Psi_{G_{x_1}}\right)(N(m_1, P_1)) = N((H_{x_n} \circ \ldots \circ H_{x_1})(m_1, P_1)) \tag{8.14}$$

with the composition conjugacy mappings

$$(H_{x_n} \circ \ldots \circ H_{x_1})(m_1, P_1) \quad = \quad \left((nP_0+P_1)^{-1}P_0 \sum_{1 \le k \le n} x_k + (nP_0+P_1)^{-1}P_1\ m_1,\ nP_0+P_1\right)$$

In a more synthetic form, we have proved that

$$\left\{\begin{array}{rcl} \Theta & \sim & \mathcal{N}(m_1, P_1) \\ (X_1, \ldots, X_n) \text{ conditional i.i.d.} \mid \Theta & \sim & \mathcal{N}(\Theta, P_0) \end{array}\right.$$

$$\Longrightarrow \quad \Theta \mid (X_1, \ldots, X_n) \ \sim \ \mathcal{N}((H_{X_n} \circ \ldots \circ H_{X_1})(m_1, P_1))$$

The class of inverse Wishart distributions is conjugate to Gaussian likelihood functions *with known mean vector*.

We recall that the distribution of a $(d \times d)$-Wishart random matrix $A \sim \mathcal{W}_d(v, V)$, with parameter $v > d$ and some $(d \times d)$ covariance matrix V, is given by

$$\mathbb{P}(A \in da) \propto |a|^{\frac{(v-d)-1}{2}} \exp\left(-\frac{1}{2} \text{ tr}\left(V^{-1} a\right)\right) da$$

where $da = \prod_{1 \leq i \leq j \leq d} da_{i,j}$ stands for the infinitesimal neighborhood of a positive definite symmetric matrix $a = (a_{i,j})_{1 \leq i,j \leq d}$. One can also show that $A \overset{law}{=} \sum_{1 \leq k \leq v} X_k X_k'$, with v independent Gaussian random vectors $X_k \sim \mathcal{N}(0, V)$. Notice that for $d = 1$ and $V = 1$ the Wishart reduces to the Chi-square distribution on $]0, \infty[$ given by

$$\mathbb{P}(A \in da) \propto a^{\frac{v}{2}-1} \exp\left(-\frac{a}{2}\right) da$$

The inverse $B = A^{-1}$ is distributed with the inverse Wishart distribution $\mathcal{IW}_d(\nu, W)$ with $W = V^{-1}$, given by

$$\mathbb{P}(B \in db) \propto |b|^{-\frac{(v+d+1)}{2}} \exp\left(-\frac{1}{2} \text{ tr}\left(W \ b^{-1}\right)\right) db$$

where $db = \prod_{1 \leq i \leq j \leq d} db_{i,j}$ stands for the infinitesimal neighborhood of a positive definite symmetric matrix $b = (b_{i,j})_{1 \leq i,j \leq d}$. We check this claim, using the fact that the Jacobian of the transformation $a \mapsto b = a^{-1}$ is given by $|\partial b / \partial a| = |a|^{-(d+1)}$. Further details on these multivariate Gaussian distributions can be found in [9]. Notice that for $d = 1$ and $(\alpha, \beta) = (v/2, W/2)$, the inverse Wishart reduces to the inverse Gamma distribution on $]0, \infty[$ given by

$$\mathbb{P}(B \in db) \propto \frac{1}{b^{\alpha+1}} \exp\left(-\frac{\beta}{b}\right) db$$

For any fixed mean vector $m_0 \in \mathbb{R}^d$ the set of prior distributions is given by

$$\forall h = (v, W) \in \mathcal{H} = (\mathbb{N}_d \times \Sigma_d) \qquad \nu_h := \mathcal{IW}_d(\nu, W) \in \mathcal{P}(\Xi) = \mathcal{P}(\Sigma_d)$$

with $\mathbb{N}_d = \{m \in \mathbb{N} \ : \ m > d\}$. and the likelihood functions are defined by the centered Gaussian densities

$$\forall x \in E = \mathbb{R}^d \qquad G_x(b) = g_{(0,b)}(x) \tag{8.15}$$

In this notation, we have

$$\Psi_{G_x}\left(\mathcal{IW}_d(v, W)\right) = \mathcal{IW}_d(H_x(v, W)) \tag{8.16}$$

with the conjugacy mapping (8.9) is given by

$$H_x(v, W) = \left(v + 1, W + xx'\right) = (v, W) + (1, xx')$$

We check this claim using the fact that $x' b^{-1} x = \text{tr}\left(xx' \ b^{-1} \right)$. The corresponding Bayesian learning model (8.5) is now given by

$$\left\{ \begin{array}{rcl} \Theta & \sim & \mathcal{IW}_d(v, W) \\ X \mid \Theta & \sim & \mathcal{N}(m_0, \Theta) \end{array} \right. \implies \Theta \mid X \ \sim \ \mathcal{IW}_d(H_X(v, W)) \tag{8.17}$$

Iterating the argument, we find that

$$\left(\Psi_{G_{x_n}} \circ \ldots \circ \Psi_{G_{x_1}}\right)\left(\mathcal{IW}_d(v, W)\right) = \mathcal{IW}_d\left((H_{x_n} \circ \ldots \circ H_{x_1})(v, W)\right) \tag{8.18}$$

with the composition conjugacy mapping

$$(H_{x_n} \circ \ldots \circ H_{x_1})(v, W) \ = \ (v, W) + (n, \sum_{1 \leq k \leq n} x_k x_k')$$

8.1.3 Exponential family of distributions

 One natural way to design conjugate priors is to consider the exponential family of distributions

$$L_{1,2}(\theta, dx) = \exp\left\{A(\theta)'T(x) - B(\theta)\right\}\ \lambda(dx) \tag{8.19}$$

with some given functions $T : x \in E \mapsto \mathbb{R}^d$, $A : \theta \in \Xi \mapsto \mathbb{R}^d$, for some $d \geq 1$, $B : \theta \in \Xi \mapsto [0, \infty[$, and some reference positive measure $\lambda(dx)$.
The measure λ is sometimes called *the base distribution* (and it is usually not important), the function $A(\theta)$ is termed the *canonical parameter* (often denoted by $\eta(\theta)$) the function $B(\theta)$ coincides with the logarithm of the normalizing constant (a.k.a. the log-partition function), and the function $T(x)$ which encapsulates the data in the conditional distribution is called *a sufficient statistic*.

In this situation, there always exists conjugate priors w.r.t. the likelihood function $G_x(\theta) = \exp\{A(\theta)'\ T(x) - B(\theta)\}$.

To see this claim, it suffices to choose a prior distribution of the form

$$\pi_1(d\theta) \propto \exp\left\{A(\theta)'\alpha - \beta B(\theta)\right\}\ \mu(d\theta) \tag{8.20}$$

for some parameters $(\alpha, \beta) \in (\mathbb{R}^d \times \mathbb{R})$ and some reference measure μ on Ξ, s.t. the above probability measure is well defined. This class of measures is often called the minimal conjugate family for the likelihood functions in (8.19).

In this situation, we have

$$\begin{aligned} \pi(d(\theta, x)) &\propto \exp\left\{A(\theta)'(T(x) + \alpha) - (1 + \beta)B(\theta)\right\}\ \mu(d\theta)\ \lambda(dx) \\ &\propto \exp\left\{A(\theta)'(T(x) + \alpha) - (1 + \beta)B(\theta)\right\}\ \mu(d\theta)\ (\pi_1 L_{1,2})(dx) \end{aligned}$$

 from which we conclude that

$$L_{2,1}(x, d\theta) \propto \exp\left\{A(\theta)'\alpha(x) - \beta(x)\ B(\theta)\right\}\ \mu(d\theta)$$

with

$$\alpha(x) = (T(x) + \alpha) \quad \text{and} \quad \beta(x) = 1 + \beta$$

In the example (8.4) discussed above, we have

$$A(\theta) = \log(\theta/(1-\theta)) \quad T(x) = x \quad B(\theta) = n\log(1-\theta) \quad \text{and} \quad \lambda(dx) = \sum_{0 \leq k \leq n} \binom{n}{k} \delta_k(dx)$$

and the prior distribution π_1 defined in (8.4) has the form (8.20) with $(\alpha, \beta) = (a-1, -(a+b)/n)$. The single observation model (8.6) corresponds to the case $n = 1$.

For multiple conditional observations with distribution (8.19), we have

$$(8.19) \Longrightarrow \mathcal{L}_{1,2}(\theta, dx) = \exp\left\{ A(\theta)' \sum_{1\leq k\leq n} T(x^k) - nB(\theta) \right\} \lambda^{\otimes n}(dx)$$

where $dx = dx^1 \times \ldots \times dx^n$ stands for an infinitesimal neighborhood of the state $x = (x^1, \ldots, x^n)$. In this situation, we have

$$(8.20) \implies \pi(d(\theta, x)) = \pi_2(dx)\ \mathcal{L}_{2,1}(x, d\theta)$$

with the posterior distribution

$$\mathcal{L}_{2,1}(x, d\theta) \propto \exp\left\{ A(\theta)'\alpha_n(x) - \beta_n(x)\ B(\theta) \right\}\ \mu(d\theta)$$

and the parameters

$$\alpha_n(x) = \sum_{1\leq k\leq n} T(x^k) + \alpha \quad \text{and} \quad \beta_n(x) = n + \beta$$

Hierarchical Bayesian models consists with introducing another level of prior distributions on the parameters (α, β).

8.2 Riemannian manifolds and Information theory

8.2.1 Nash embedding theorem

In differential geometry, a (smooth) Riemannian manifold $(\mathcal{S}, g)$ is a real state space $\mathcal{S}$ equipped with a smooth inner product g on the tangent space $T(S)$; that is, for any $\theta \in \mathcal{S}$, and any vector fields $\theta \mapsto V_i(\theta) \in T_\theta(S)$ the mapping

$$\theta \mapsto \langle V_1(\theta), V_2(\theta) \rangle_{g(\theta)}$$

is a smooth function. This geometric Riemannian structure allows to define various geometric notions such as angles, lengths of curves, volumes, curvature, gradients of functions and divergence of vector fields.

The Nash embedding theorem state that every Riemannian manifold with dimension p can be (locally) isometrically embedded into some Euclidean ambient space with sufficiently high r-dimension (but $r \leq 2p + 1$). The isometric embedding problem amounts to find some some function $\psi : \theta \in \mathcal{S} \mapsto \psi(\theta) \in \mathbb{R}^r$ such that

$$g_{i,j}(\theta) := \langle \partial_{\theta_i}\psi, \partial_{\theta_j}\psi \rangle = \sum_{1\leq k\leq r} \partial_{\theta_i}\psi^k(\theta)\partial_{\theta_j}\psi^k(\theta)$$

8.2.2 Distribution manifolds

When $\mathcal{S} = S_\psi = \phi(S)$ is the parameter space of a given manifold S discussed in (6.36) the natural Riemannian inner product is given by the matrix field (6.39).

The space of discrete distributions $S := \mathcal{P}(E)$ on a finite set $E = \{1, \ldots, r\}$ is represented by the $p = (r - 1)$-dimensional simplex

$$\text{Simplex}(p) = \{z = (z_i)_{1\leq i\leq r} \in \mathbb{R}^r_+ \ :\ \varphi(z) := \sum_{1\leq i\leq r} z_i - 1\}$$

The tangent space $T_z(S)$ at each point z is given by

$$\partial\varphi = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Rightarrow T_z(S) = \Big\{ W(z) \in \mathbb{R}^r \ : \ z^k = 0 \Rightarrow W^k(z) = 0 \quad \text{and} \quad \langle W(z), \partial\varphi(z) \rangle = \textstyle\sum_{1 \leq k \leq r} W^k(z) = 0 \Big\}$$

The Fisher information metric on $T_z(S)$ is defined by the inner product

$$\forall W_1(z), W_2(z) \in T_z(S) \quad \langle W_1(z), W_2(z) \rangle_{h(z)} = \sum_{1 \leq k \leq r} \frac{W_1^k(z)}{z_k} \frac{W_2^k(z)}{z_k} \ z_k$$

For instance, for $r = 3$ we have

$$T_z(S) := \text{Vect} \left(e_1(z) := \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, e_2(z) := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right)$$

In this case, for any $z = (z_k)_{1 \leq k \leq 3}$ s.t. $z_k > 0$ for any $k = 1, 2, 3$, we have

$$\begin{aligned} h_{1,1}(z) &= \langle e_1(z), e_1(z) \rangle_{h(z)} = \frac{1}{z_1} + \frac{1}{z_3} \quad h_{1,2}(z) = h_{2,1}(z) = \langle e_1(z), e_2(z) \rangle_{h(z)} = \frac{1}{z_3} \\ h_{2,2}(z) &= \langle e_2(z), e_2(z) \rangle_{h(z)} = \frac{1}{z_2} + \frac{1}{z_3} \end{aligned}$$

More generally, let E be some measurable space equipped with some reference measure λ. The tangent space $T_\mu \left(\mathcal{P}^\lambda(E) \right)$ of the set of probability measures

$$\mathcal{P}^\lambda(E) := \{ \mu \in \mathcal{P}(E) \ : \ \mu \ll \lambda \ \} \ni \mu$$

given by

$$T_\mu \left(\mathcal{P}^\lambda(E) \right) = \left\{ \nu \in \mathcal{P}(E) \ : \ \nu \ll \mu \text{ s.t. } \int \left(\frac{d\nu}{d\mu} \right)^2 \ d\mu < \infty \quad \text{and} \quad \nu(1) = 0 \right\}$$

is equipped with the Fisher inner product

$$\forall W_1(\mu), W_2(\mu) \in T_\mu(\mathcal{P}^\lambda(E)) \quad \langle W_1(\mu), W_2(\mu) \rangle_{h(\mu)} := \int \frac{dW_1(\mu)}{d\mu} \frac{dW_2(\mu)}{d\mu} \ d\mu \tag{8.21}$$

8.2.3 Bayesian statistical manifolds

Riemannian manifolds also arise in a natural way in Bayesian statistics and Information theory. To describe with some precision these statistical models, we let

$$\mu_\theta(dy) := P_\theta(y) \ \lambda(dy) \tag{8.22}$$

be a collection of distributions on some state space E, equipped with some reference measure $\lambda(dy)$, and indexed by some parameter θ on some space $\mathcal{S} \subset \mathbb{R}^p$ of dimension p. We assume that $\mathcal{S}$ is equipped with some probability measure of the form

$$\mu'(d\theta) = P'(\theta)\ \lambda'(d\theta)$$

where $\lambda'(d\theta)$ stands for some reference measure on $\mathcal{S}$. The probability measure μ' can be seen as the prior distribution of some unknown random parameter Θ. Given $\Theta = \theta$, $\mu_\theta(dy)$ stands for the distribution of some partial and noisy random observation Y of the parameter Θ. In this interpretation, the function

$$P(\theta, y) = P'(\theta)\ P_\theta(y)$$

represents the density of the random variable (Θ, Y) w.r.t. the reference measure $\lambda \otimes \lambda'$ on $\mathcal{S} \times E$. We consider the parametrization mapping

$$\psi\ :\ \theta \in \mathcal{S} \mapsto \psi(\theta) = \mu_\theta \in \mathcal{P}_{\mathcal{S}}(E) = \{\mu_\theta \in \mathcal{P}(E)\ :\ \theta \in \mathcal{S}\ \} \subset \mathcal{P}^\lambda(E)$$

and we equip $\mathcal{P}_{\mathcal{S}}(E)$ with the Fisher metric (8.21) induced by $\mathcal{P}^\lambda(E)$. Notice that for any $1 \leq i \leq p$ we have

$$(\partial_{\theta_i}\psi)(\theta) = \partial_{\theta_i}\mu_\theta$$

with the signed measure

$$\partial_{\theta_i}\mu_\theta(dy) := \partial_{\theta_i}P_\theta(y)\ \lambda(dy)$$

on E with null mass

$$\int P_\theta(y)\ \lambda(dy) = 1 \Rightarrow \forall 1 \leq i \leq p \quad \int \partial_{\theta_i}P_\theta(y)\ \lambda(dy) = 0$$

The tangent space

$$T_\theta\left(\mathcal{S}\right) = \mathrm{Vect}\left(e_i,\ i = 1, \ldots, p\right) \quad \text{with} \quad e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \longleftarrow\ i - \text{th coordinate}$$

is mapped on the tangent space $T_{\mu_\theta}\left(\mathcal{P}_{\mathcal{S}}(E)\right)$ using the push forward mapping

$$(d\psi)_\theta\ :\ V(\theta) = \sum_{1 \leq i \leq p} V^i(\theta)\ e_i(\theta) \in T_\theta\left(\mathcal{S}\right) \mapsto (d\psi)_\theta\left(V(\theta)\right) = \sum_{1 \leq i \leq p} V^i(\theta)\ (\partial_{\theta_i}\psi)(\theta) \in T_{\mu_\theta}\left(\mathcal{P}_{\mathcal{S}}(E)\right)$$

The Fisher information metric g on the parameter space $\mathcal{S}$ induced by the metric h on $\mathcal{P}_{\mathcal{S}}(E)$ is defined for any $1 \leq i, j \leq p$ by

$$\begin{aligned} g_{i,j}(\theta) &:= \left\langle (\partial_{\theta_i}\psi)(\theta), (\partial_{\theta_j}\psi)(\theta) \right\rangle_{h(\mu_\theta)} = \int \frac{\partial_{\theta_i}P_\theta(y)}{P_\theta(y)}\ \frac{\partial_{\theta_j}P_\theta(y)}{P_\theta(y)}\ P_\theta(y)\ \lambda(dy) \\ &= \int \partial_{\theta_i}\log P_\theta(y)\ \partial_{\theta_j}\log P_\theta(y)\ P_\theta(y)\ \lambda(dy) \\ &= \mathbb{E}\left(\partial_{\theta_i}\log P_\Theta(Y)\ \partial_{\theta_j}\log P_\Theta(Y)\ |\ \Theta = \theta\ \right) \end{aligned}$$

The Fisher metric can alternatively be defined by

$$\begin{aligned}\int P_\theta(y)\ \lambda(dy) = 1 \quad &\Rightarrow \quad \int \partial_{\theta_i} \log P_\theta(y)\ P_\theta(y)\ \lambda(dy) = 0 \\ &\Rightarrow \quad g_{i,j}(\theta) \ = \ \int \partial_{\theta_i} \log P_\theta(y)\ \partial_{\theta_j} \log P_\theta(y)\ P_\theta(y)\ \lambda(dy) \\ &= \ -\int \left(\partial_{\theta_j,\theta_i} \log P_\theta(y)\right)\ P_\theta(y)\ \lambda(dy) \\ &= \ -\mathbb{E}\left(\partial_{\theta_j,\theta_i} \log P_\Theta(Y) \mid \Theta = \theta\ \right)\end{aligned}$$

We end this section with a connection between the Fisher metric and the relative Boltzmann entropy. We fix some parameter $\theta^\star \in \mathcal{S}$ and we consider the Boltzmann entropy

$$\mathcal{B}_{\theta^\star}(\theta) = \mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right) = -\int \ \log \frac{P_\theta(y)}{P_{\theta^\star}(y)}\ P_{\theta^\star}(y)\ \lambda(dy)$$

We have

$$\begin{aligned}\partial_{\theta_i} \log P_\theta(y) \ &= \ \frac{1}{P_\theta(y)}\ \partial_{\theta_i} P_\theta(y) \\ \partial_{\theta_j,\theta_i} \log P_\theta(y) \ &= \ -\frac{1}{P_\theta(y)^2}\ \partial_{\theta_j} P_\theta(y) \partial_{\theta_i} P_\theta(y) + \frac{1}{P_\theta(y)}\ \partial_{\theta_j,\theta_i} P_\theta(y)\end{aligned}$$

from which we conclude that

$$\left(\partial_{\theta_i} \mathcal{B}_{\theta^\star}\right)(\theta^\star) = -\int \ \frac{1}{P_{\theta^\star}(y)}\ \partial_{\theta_i} P_{\theta^\star}(y)\ P_{\theta^\star}(y)\ \lambda(dy) = -\int \ \partial_{\theta_i} P_{\theta^\star}(y)\ \ \lambda(dy) = 0$$

and

$$\begin{aligned}\left(\partial_{\theta_j,\theta_i} \mathcal{B}_{\theta^\star}\right)(\theta^\star) \ &= \ \int \ \frac{1}{P_{\theta^\star}(y)^2}\ \partial_{\theta_j} P_{\theta^\star}(y) \partial_{\theta_i} P_{\theta^\star}(y)\ P_{\theta^\star}(y) \lambda(dy) - \int \ \partial_{\theta_j,\theta_i} P_{\theta^\star}(y) \lambda(dy) \\ &= \ \int \ \frac{\partial_{\theta_j} P_{\theta^\star}(y)}{P_{\theta^\star}(y)}\ \frac{\partial_{\theta_j} P_{\theta^\star}(y)}{P_{\theta^\star}(y)}\ P_{\theta^\star}(y) \lambda(dy) \ = \ g_{i,j}(\theta^\star)\end{aligned}$$

This shows that

$$\begin{aligned}\mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right) \ &= \ \frac{1}{2} \sum_{1 \leq i,j \leq p} g_{i,j}(\theta^\star)\ (\theta_i - \theta_i^\star)(\theta_j - \theta_j^\star) + \mathrm{O}(\|(\theta - \theta^\star)\|^3) \\ &= \ \frac{1}{2}\ (\theta - \theta^\star)^T g(\theta^\star)(\theta - \theta^\star) + \mathrm{O}(\|(\theta - \theta^\star)\|^3)\end{aligned}$$

The above formula shows that the Fisher matrix $g(\theta)$ encapsulate the infinitesimal changes of the model distribution μ_θ w.r.t. an infinitesimal fluctuation of the model parameter θ.

8.2.4 The Cramer-Rao lower bound

Suppose we are given an unbias estimate $\widehat{\Theta} = (\varphi^i(Y))_{1 \leq i \leq p}$ of the parameter $\theta = (\theta^i)_{1 \leq i \leq p}$ associated with an observation r.v. Y with distribution (8.22); that is we have that

$$\forall 1 \leq i \leq p \qquad \mathbb{E}\left(\varphi^i(Y)\right) = \theta^i$$

The score function is defined by the gradient function

$$\mathbf{Score}_\theta^i(Y) := \partial_{\theta_i} \log P_\theta(Y)$$

Recalling that

$$\mathbb{E}\left(\mathbf{Score}_\theta^i(Y)\right) = 0$$

using Cauchy-Schwartz inequality, we find that

$$\begin{aligned}
\mathbb{E}\left(\left[\mathbf{Score}_\theta^i(Y)\right]^2\right)^{1/2} \times \mathrm{Var}(\varphi^j(Y))^{1/2} &\geq \mathbb{E}\left(\left[\mathbf{Score}_\theta^i(Y) - \mathbb{E}\left(\mathrm{Score}_\theta^i(Y)\right)\right]\left[\varphi^j(Y) - \mathbb{E}\left(\varphi^j(Y)\right)\right]\right) \\
&= \mathbb{E}\left(\mathbf{Score}_\theta^i(Y)\varphi^j(Y)\right) \\
&= \int \varphi^j(y)\ \partial_{\theta_i} \log P_\theta(y)\ P_\theta(y)\ \lambda(dy) \\
&= \int \varphi^j(y)\ \partial_{\theta_i} P_\theta(y)\ \lambda(dy) = \partial_{\theta_i}\mathbb{E}(\varphi^j(Y)) = \partial_{\theta_i}\theta^j = 1_{i=j}
\end{aligned}$$

This implies that

$$\mathrm{Var}(\varphi^j(Y)) \geq 1/g_{j,j}(\theta)$$

8.2.5 Some illustrations

Boltzmann-Gibbs measures

We consider a collection of Boltzmann-Gibbs measures associated with some potential function V on some state space E, and indexed by some real valued parameter θ:

$$\mu_\theta(dy) = \frac{1}{\mathcal{Z}_\theta}\ e^{-\theta V(y)}\ \lambda(dy)$$

In this situation, we have

$$\begin{aligned}
\partial_\theta P_\theta(y) &= -\frac{1}{\mathcal{Z}_\theta^2}\ \partial_\theta(\mathcal{Z}_\theta)\ e^{-\theta V(y)} + \frac{1}{\mathcal{Z}_\theta}\ \partial_\theta(e^{-\theta V(y)}) \\
&= (\mu_\theta(V) - V(y))\ P_\theta(y) \Longrightarrow \partial_\theta \log P_\theta = \mu_\theta(V) - V)
\end{aligned}$$

from which we conclude that

$$g(\theta) = g_{1,1}(\theta) = \int\ [\mu_\theta(V) - V(y)]^2\ \mu_\theta(dy) = \mu_\theta(V^2) - \mu_\theta(V)^2$$

Multivariate normal distributions

We consider the collection of distributions μ_θ index by some parameter $\theta \in \mathcal{S} \subset \mathbb{R}^p$ and given by

$$\mu_\theta(dy) = \frac{1}{\sqrt{2\pi}^{d^Y}\ \sqrt{\det(C(\theta))}}\ \exp\left(-\frac{1}{2}\ (y - m(\theta))^T C(\theta)^{-1}(y - m(\theta))\right)\ dy$$

where $dy = \prod_{1\leq i\leq d^Y} dy_i$ stands for an infinitesimal neighborhood of the point $y = (y_i)_{1\leq i\leq d^Y} \in \mathbb{R}^{d^Y}$.

In this situation, we have

$$\begin{aligned} g_{i,j}(\theta^\star) &\overset{\theta=\theta^\star}{=} \partial_{\theta_i,\theta_j}\mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right) \\ &\overset{\theta=\theta^\star}{=} \partial_{\theta_i} m(\theta^\star)^T C(\theta^\star)^{-1}\partial_{\theta_j} m(\theta^\star) + \frac{1}{2}\,\mathrm{tr}\left(\partial_{\theta_j} C(\theta^\star)\; C(\theta^\star)^{-1}\left(\partial_{\theta_i} C(\theta^\star)\right) C(\theta^\star)^{-1}\right) \end{aligned} \tag{8.23}$$

In particular, for $d^Y = 1$, and $p = 2$ with $m(\theta) = \theta_1 \in \mathbb{R}$ and $C(\theta) = \theta_2 \in]0, \infty[$ we have

$$\begin{aligned} g_{1,1}(\theta) &= \frac{1}{\theta_2} \\ g_{1,2}(\theta) &= g_{2,1}(\theta) = 0 \quad \text{and} \quad g_{2,2}(\theta) = \frac{1}{2\theta_2^2} \end{aligned}$$

The corresponding Riemannian gradient compensates the fact that a infinitesimal change of parameter in a Gaussian model μ_θ with small variance θ_2 has more pronounced effects:

$$(\nabla_g f)(\theta) = \theta_2\; \partial_{\theta_1} f + 2\theta_2^2\; \partial_{\theta_2} f$$

To check (8.23), we observe that

$$\begin{aligned} \mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right) &= \frac{1}{2}\Big[\log\det\left(C(\theta^\star)^{-1}C(\theta)\right) \\ &\quad + \int \left\{\left((y-m(\theta))^T C(\theta)^{-1}(y-m(\theta))\right) - (y-m(\theta^\star))^T C(\theta^\star)^{-1}(y-m(\theta^\star))\right\}\mu_{\theta^\star}(dy)\Big] \\ &= \tfrac{1}{2}\Big[\log\det\left(C(\theta^\star)^{-1}C(\theta)\right) \\ &\quad + \mathbb{E}\left\{\left((Y-m(\theta))^T C(\theta)^{-1}(Y-m(\theta))\right) - (Y-m(\theta^\star))^T C(\theta^\star)^{-1}(Y-m(\theta^\star))\right\}\Big] \end{aligned}$$

with

$$\begin{aligned} Y &= m(\theta^\star) + C(\theta^\star)^{1/2}\; Z \quad \text{where} \quad Z \sim N(0, Id_{d^Y \times d^Y}) \\ \mathbb{E}\left(ZAZ^T\right) &= \sum_{1\le i,j\le d^Y} \mathbb{E}(Z^i A_{i,j}\; Z^j) = \sum_{1\le i,j\le d^Y} A_{i,i} = \mathrm{tr}(A) \end{aligned}$$

We recall that for any invertible symmetric positive definite matrix A, the square root $A^{1/2} = \left(U\sqrt{D}U^T, \text{ for some orthonormal diagonalizing matrix } U \text{ s.t. } A = UDU^T \text{ with } D \text{ diagonal}\right)$ is symmetric and invertible, and we have

$$A^{-1/2}A^{1/2} = Id \qquad AA^{1/2} = A^{1/2}A \quad \text{and} \quad A^{1/2}A^{-1}A^{1/2} = Id$$

We also recall that for any couple of matrices A and B, we have

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

so that

$$\mathrm{tr}(A^{1/2}(BA^{1/2})) = \mathrm{tr}(BA) = \mathrm{tr}(AB)$$

Using these formula, it is readily checked that

$$\mathbb{E}\left((Y-m(\theta^\star))^T C(\theta^\star)^{-1}(Y-m(\theta^\star))\right) = \mathbb{E}\left(Z^T\; C(\theta^\star)^{1/2}\; C(\theta^\star)^{-1}C(\theta^\star)^{1/2}\; Z\right) = \mathbb{E}\left(Z^T Z\right) = d^Y$$

and

$$\begin{aligned}
&\mathbb{E}\left((Y-m(\theta))^T C(\theta)^{-1}(Y-m(\theta))\right)\\
&= \mathbb{E}\left(\left[(m(\theta^\star)-m(\theta))+C(\theta^\star)^{1/2}Z\right]^T C(\theta)^{-1}\left[(m(\theta^\star)-m(\theta))+C(\theta^\star)^{1/2}Z\right]\right)\\
&= (m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta)) + \mathbb{E}\left(Z^T C(\theta^\star)^{1/2}C(\theta)^{-1}C(\theta^\star)^{1/2}Z\right)\\
&= (m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta)) + \mathrm{tr}\left(C(\theta^\star)^{1/2}C(\theta)^{-1}C(\theta^\star)^{1/2}\right)\\
&= (m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta)) + \mathrm{tr}\left(C(\theta^\star)C(\theta)^{-1}\right)
\end{aligned}$$

We conclude that

$$\begin{aligned}
&2\,\mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right)\\
&= \log\det\left(C(\theta^\star)^{-1}C(\theta)\right) + (m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta)) + \left[\mathrm{tr}\left(C(\theta^\star)C(\theta)^{-1}\right) - d^Y\right]
\end{aligned}$$

For any index $1 \le i \le p$, we have

$$\begin{aligned}
\partial_{\theta_i}\log\det\left(C(\theta^\star)^{-1}C(\theta)\right) &= \frac{1}{\det\left(C(\theta^\star)^{-1}C(\theta)\right)}\,\partial_{\theta_i}\log\det\left(C(\theta^\star)^{-1}C(\theta)\right)\\
&= \mathrm{tr}\left(\left(C(\theta^\star)^{-1}C(\theta)\right)^{-1}\partial_{\theta_i}C(\theta^\star)^{-1}C(\theta)\right)\\
&= \mathrm{tr}\left(C(\theta)^{-1}\partial_{\theta_i}C(\theta)\right)
\end{aligned}$$

so that

$$\begin{aligned}
&\partial_{\theta_i}\left(\log\det\left(C(\theta^\star)^{-1}C(\theta)\right) + \left[\mathrm{tr}\left(C(\theta^\star)C(\theta)^{-1}\right) - d^Y\right]\right)\\
&= \mathrm{tr}\left(C(\theta)^{-1}\partial_{\theta_i}C(\theta) - C(\theta^\star)C(\theta)^{-1}\left(\partial_{\theta_i}C(\theta)\right)C(\theta)^{-1}\right) \overset{\theta=\theta^\star}{=} 0
\end{aligned}$$

The second term in the trace formula comes from the fact that

$$\partial_\epsilon A(\epsilon)^{-1} = -A(\epsilon)^{-1}\ \left(\partial_\epsilon A(\epsilon)\right)\ A(\epsilon)^{-1}$$

for any smooth functional $\epsilon \mapsto A(\epsilon)$ in the space of invertible matrices. We check this claim using the fact that

$$\begin{aligned}
\partial_\epsilon \sum_j A_{i,j}(\epsilon)A^{j,k}(\epsilon) = 0 \quad &\Rightarrow \quad \sum_j A_{i,j}(\epsilon)\partial_\epsilon A^{j,k}(\epsilon) = -\sum_j \partial_\epsilon A_{i,j}(\epsilon)A^{j,k}(\epsilon)\\
&= \quad \sum_{i,j} A^{l,i}(\epsilon)A_{i,j}(\epsilon)\partial_\epsilon A^{j,k}(\epsilon) = -\sum_{i,j} A^{l,i}(\epsilon)\partial_\epsilon A_{i,j}(\epsilon)A^{j,k}(\epsilon)
\end{aligned}$$

where $A(\epsilon)^{-1} = \left(A^{i,j}(\epsilon)\right)_{i,j}$ and $A(\epsilon) = \left(A_{i,j}(\epsilon)\right)_{i,j}$.

We also have that

$$\begin{aligned}
&\partial_{\theta_i,\theta_j}\left(\log\det\left(C(\theta^\star)^{-1}C(\theta)\right) + \left[\mathrm{tr}\left(C(\theta^\star)C(\theta)^{-1}\right) - d^Y\right]\right)\\
&= \mathrm{tr}\left\{\left[\left(\partial_{\theta_j}C(\theta)^{-1}\right)\ \partial_{\theta_i}C(\theta) + C(\theta)^{-1}\partial_{\theta_i,\theta_j}C(\theta)\right]\right.\\
&\left. -C(\theta^\star)\left[\left(\partial_{\theta_j}C(\theta)^{-1}\right)\left(\partial_{\theta_i}C(\theta)\right)C(\theta)^{-1} + C(\theta)^{-1}\left(\partial_{\theta_i,\theta_j}C(\theta)\right)C(\theta)^{-1} + C(\theta)^{-1}\left(\partial_{\theta_i}C(\theta)\right)\ \partial_{\theta_j}C(\theta)^{-1}\right]\right\}\\
&\overset{\theta=\theta^\star}{=} -\mathrm{tr}\left(C(\theta^\star)\left(\partial_{\theta_j}C(\theta^\star)^{-1}\right)\left(\partial_{\theta_i}C(\theta^\star)\right)C(\theta^\star)^{-1}\right) = \mathrm{tr}\left(\partial_{\theta_j}C(\theta^\star)\ C(\theta^\star)^{-1}\left(\partial_{\theta_i}C(\theta^\star)\right)C(\theta^\star)^{-1}\right)
\end{aligned}$$

In much the same way, we have

$$\partial_{\theta_i}\left((m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta))\right)$$

$$= -2\partial_{\theta_i} m(\theta)^T C(\theta)^{-1}(m(\theta^\star)-m(\theta)) + \left((m(\theta^\star)-m(\theta))^T \partial_{\theta_i} C(\theta)^{-1}(m(\theta^\star)-m(\theta))\right) \stackrel{\theta=\theta^\star}{=} 0$$

and therefore

$$\partial_{\theta_i,\theta_j}\left((m(\theta^\star)-m(\theta))^T C(\theta)^{-1}(m(\theta^\star)-m(\theta))\right)$$

$$= 2\ \partial_{\theta_i} m(\theta)^T C(\theta)^{-1}\partial_{\theta_j} m(\theta)$$

$$-2\ \Big[\left(\partial_{\theta_i\theta_j} m(\theta)\right)^T C(\theta)^{-1} + \left(\partial_{\theta_i} m(\theta)\right)^T \partial_{\theta_j} C(\theta)^{-1} + \partial_{\theta_j} m(\theta)^T \partial_{\theta_i} C(\theta)^{-1}$$

$$-\frac{1}{2}(m(\theta^\star)-m(\theta))^T \partial_{\theta_i,\theta_j} C(\theta)^{-1}\Big]\,(m(\theta^\star)-m(\theta))$$

$$\stackrel{\theta=\theta^\star}{=} 2\ \partial_{\theta_i} m(\theta^\star)^T C(\theta^\star)^{-1}\partial_{\theta_j} m(\theta^\star)$$

We conclude that

$$\partial_{\theta_j}\mathrm{Ent}\left(\mu_{\theta^\star} \mid \mu_\theta\right) \stackrel{\theta=\theta^\star}{=} 0$$

This ends the proof of the desired formula.

8.3 Signal processing and filtering

This section is dedicated to linear-Gaussian filtering models and the derivation of the traditional forward-backward Kalman filters. The origins of these optimal filters starts with the seminal article by R. E. Kalman [370] and the earlier pioneering works by R. L. Startonovich [554, 555, 556, 557].

The first historical application of the Kalman filter was developed by S. F. Schmidt [528], in the Apollo program of NASA Ames Research Center, to solve nonlinear navigation equations of the manned lunar mission. The Kalman filter has been applied in the design of a variety of defense navigation and guidance systems, including ballistic submarines and the U.S. Navy's Tomahawk as well as in the U.S. Air Force's air launched cruise missiles. It is also used in the NASA Space Shuttle, as well as in the attitude control systems of the International Space Station.

Originally developed for spacecraft navigation systems, the Kalman filter and its extended version are one of the most commonly and routinely used tools to remove noise from partially observed sequences of random variables. Its range of application has been extended to almost every scientific discipline, including in financial mathematics, econometrics, and computational biology, as well as in Bayesian statistics and in various branches of engineering sciences.

8.3.1 Forward Kalman filters

We consider a $\mathbb{R}^{p+q}$-valued Markov chain (X_n, Y_n) defined by the recursive relations

$$\left\{\begin{array}{rcll} X_n &=& A_n\ X_{n-1} + a_n + B_n W_n\,, & n \geq 1 \\ Y_n &=& C_n\ X_n + c_n + D_n\ V_n\,, & n \geq 0 \end{array}\right. \tag{8.24}$$

for some $\mathbb{R}^{d_w}$ and $\mathbb{R}^{d_v}$-valued independent random sequences W_n and V_n, independent of X_0, some matrices A_n, B_n, C_n, D_n with appropriate dimensions and finally some $(p+q)$-dimensional vector (a_n, c_n). We further assume that W_n and V_n centered Gaussian random sequences with covariance

matrices R_n^v, R_n^w and X_0 is a Gaussian random variable in $\mathbb{R}^p$ with a mean and covariance matrix denoted by

$$\widehat{X}_0^- = \mathbb{E}(X_0) \quad \text{and} \quad \widehat{P}_0^- = \mathbb{E}((X_0 - \mathbb{E}(X_0))\ (X_0 - \mathbb{E}(X_0))')$$

The one-step predictors and the optimal filters are given by

$$\begin{aligned} \eta_n &= \text{Law}(X_n \mid (Y_0, \ldots, Y_{n-1})) = \mathcal{N}(\widehat{X}_n^-, P_n^-) \\ \widehat{\eta}_n &= \text{Law}(X_n \mid (Y_0, \ldots, Y_{n-1}, Y_n)) = \mathcal{N}(\widehat{X}_n, P_n) \end{aligned} \tag{8.25}$$

In the above display, $\mathcal{N}(.,.)$ stands for the Gaussian distributions discussed in (8.11).

For the linear-Gaussian models discussed above, the synthesis of the conditional mean and covariance matrices is carried out using the traditional Kalman-Bucy recursive updating-prediction equations

$$\left(\widehat{X}_n^-, P_n^-\right) \xrightarrow{\text{updating}} \left(\widehat{X}_n, P_n\right) \xrightarrow{\text{prediction}} \left(\widehat{X}_{n+1}^-, P_{n+1}^-\right) \tag{8.26}$$

To derive with some precision these recursions, we let M_n be the Gaussian transition on $\mathbb{R}^p$ defined by

$$M_n(x, dx') = g_{(m_n^X(x), R_n)}(x')\ dx' \quad \text{with} \quad m_n^X(x) = A_n\ x + a_n \quad \text{and} \quad R_n := B_n R_n^w B_n' \tag{8.27}$$

We also consider the likelihood functions

$$g_{(Y_n - c_n, Q_n)}(C_n x) = g_{(m_n^Y(x), Q_n)}(Y_n) \quad \text{with} \quad m_n^Y(x) = C_n\ x + c_n \quad \text{and} \quad Q_n := D_n R_n^v D_n' \tag{8.28}$$

with the Gaussian densities $g_{(m,R)}(x)$ associated with a mean and covariance matrix (m, R) introduced in (8.11).

To find the prediction step we simply observe that

$$\begin{aligned} \widehat{X}_n^- &= \mathbb{E}(A_n X_{n-1} + a_n + B_n W_n \mid\ (Y_0, \ldots, Y_{n-1})) = A_n\ \widehat{X}_{n-1} + a_n \\ P_n^- &= \mathbb{E}\left(\left(A_n(X_{n-1} - \widehat{X}_{n-1}) + B_n W_n\right)\left(A_n(X_{n-1} - \widehat{X}_{n-1}) + B_n W_n\right)'\right) = A_n P_{n-1} A_n' + R_n \end{aligned}$$

This yields

$$\begin{aligned} \widehat{X}_n^- &= m_n^X\left(\widehat{X}_{n-1}\right) = A_n\ \widehat{X}_{n-1} + a_n \\ P_n^- &= A_n P_{n-1} A_n' + R_n \end{aligned}$$

The updating step is partly based on the fact that the $\mathcal{Y}$-martingale difference $(\widehat{X}_n - \widehat{X}_n^-)$ has the representation property with respect to the innovation process; that is, we have

$$\widehat{X}_n - \widehat{X}_n^- = \mathbf{Gain}_n\ (Y_n - \widehat{Y}_n^-) \quad \text{with} \quad \widehat{Y}_n^- = \mathbb{E}(Y_n|Y_0,\ldots,Y_{n-1}) = m_n^Y\left(\widehat{X}_n^-\right)$$

for some gain matrix $\mathbf{Gain}_n$. Since we have $\mathbb{E}((X_n - \widehat{X}_n)(Y_n - \widehat{Y}_n^-)') = 0$, and

$$(Y_n - \widehat{Y}_n^-) = C_n(X_n - \widehat{X}_n^-) + D_nV_n$$

we find that

$$\mathbb{E}((X_n - \widehat{X}_n^-)(Y_n - \widehat{Y}_n^-)') = \mathbf{Gain}_n\ \mathbb{E}((Y_n - \widehat{Y}_n^-)(Y_n - \widehat{Y}_n^-)')$$

We conclude that

$$\mathbf{Gain}_n = P_n^- C_n'(C_nP_n^-C_n' + Q_n)^{-1}$$

Finally, using the decomposition

$$X_n - \widehat{X}_n = (X_n - \widehat{X}_n^-) + (\widehat{X}_n^- - \widehat{X}_n)$$

and by symmetry argument, we conclude that

$$\begin{aligned} P_n &= P_n^- - \mathbb{E}((\widehat{X}_n^- - \widehat{X}_n)(\widehat{X}_n^- - \widehat{X}_n)') \\ &= P_n^- - \mathbf{Gain}_n\ \mathbb{E}((Y_n - \widehat{Y}_n^-)(Y_n - \widehat{Y}_n^-)')\mathbf{Gain}_n' = P_n^- - \mathbf{Gain}_nC_nP_n^- \end{aligned}$$

In summary, we have proved the following updating formula.

For any $(r, R) \in (\mathbb{R}^p \times \mathbb{R}^{p\times p})$ and $n \geq 0$ the updating transition is given by the Boltzmann-Gibbs transformation

$$\Psi_{g_{(Y_n - c_n, Q_n)}(C_n\bullet)}(\mathcal{N}(r,R)) = \mathcal{N}\left(\widehat{m}_{Y_n,n}(r,R), \widehat{\Sigma}_n(R)\right) \tag{8.29}$$

with the functionals

$$\begin{aligned} \widehat{m}_{Y_n,n}(r,R) &= r + RC_n'\Sigma_n(R)^{-1}\ (Y_n - (C_nr + c_n)) \\ \widehat{\Sigma}_n(R) &= (I - RC_n'\Sigma_n(R)^{-1}C_n)R \quad \text{and} \quad \Sigma_n(R) := C_nRC_n' + Q_n \end{aligned}$$

It is also useful to observe that

$$\text{Law}(Y_n \mid Y_0,\ldots,Y_{n-1}) = \mathcal{N}\left(C_n\widehat{X}_n^- + c_n, \Sigma_n(P_n^-)\right)$$

We prove this claim using the fact that, given $(Y_0,\ldots,Y_{n-1})$, the current observation takes the form

$$Y_n = C_n\tilde{X}_n + D_nV_n \quad \text{with} \quad \text{Law}\left(\tilde{X}_n \mid Y_0,\ldots,Y_{n-1}\right) := \mathcal{N}(\widehat{X}_n^-, P_n^-)$$

The density $p_n(y_0,\ldots,y_n)$ of the sequence of observation $(Y_0,\ldots,Y_n)$ evaluated at the random observation path $(Y_0,\ldots,Y_n)$ is given by

$$p_n(Y_0,\ldots,Y_n) = \prod_{k=0}^{n} g_{(C_k\widehat{X}_k^- + c_k, \Sigma_k(P_k^-))}(Y_k) \tag{8.30}$$

In Bayesian inference literature, this formula is sometimes written in the following form

$$\begin{aligned} p_n(Y_0,\ldots,Y_n) &= p_{n-1,n}(Y_n\ |Y_0,\ldots,Y_{n-1}) \times p_{n-1}(Y_0,\ldots,Y_{n-1}) \\ &= \prod_{k=0}^{n} p_{k-1,k}(Y_k\ |Y_0,\ldots,Y_{k-1}) \end{aligned}$$

where $p_{n-1,n}(y_n\ |y_0,\ldots,y_{n-1})$ stands for the density (w.r.t. the Lebesgue measure dy_n on $\mathbb{R}^q$) of the conditional distribution of the random variable Y_n given the observations $Y_p = y_p$, for $0 \leq p < n$.

8.3.2 Backward Kalman smoother

With a slight abuse of the notation, we denote by

$$p((x_0,\ldots,x_n)\ |\ (y_0,\ldots,y_{n-1}))$$

the density (w.r.t. the Lebesgue measure $dx_0 \times \ldots \times dx_n$ on $(\mathbb{R}^p)^{n+1}$) of the conditional distribution of the random variable $(X_0,\ldots,X_n)$ given the observations $Y_p = y_p$, for $0 \leq p < n$. We also denote by

$$p(x_k\ |\ x_{k+1},(y_0,\ldots,y_k))$$

the density (w.r.t. the Lebesgue measure dx_k on $\mathbb{R}^p$) of the conditional distribution of the random variable X_k given the observations $Y_p = y_p$, for $0 \leq p \leq k$, and the random state $X_{k+1} = x_{k+1}$.

In these Bayesian notation, we have the conditional density formulae

$$\begin{aligned} &p((x_0,\ldots,x_n)\ |\ (y_0,\ldots,y_{n-1})) \\ &= p(x_n\ |\ (y_0,\ldots,y_{n-1}))\ p(x_{n-1}\ |\ x_n,(y_0,\ldots,y_{n-1})) \\ &\quad \times p(x_{n-2}\ |\ x_{n-1},(y_0,\ldots,y_{n-2}))\ldots p(x_1\ |\ x_2,(y_0,y_1))\ p(x_0\ |\ x_1,y_0) \end{aligned} \tag{8.31}$$

 This shows that

$$\mathbb{P}\left((X_0,\ldots,X_n) \in d(x_0,\ldots,x_n)\ |\ Y_p = y_p,\ p < n\right) = \eta_n(dx_n) \prod_{1\leq k\leq n} \mathbb{M}_{k,\eta_{k-1}}(x_k,dx_{k-1}) \tag{8.32}$$

with the Markov transitions

$$\mathbb{M}_{k,\eta_{k-1}}(x_k,dx_{k-1}) = p(x_{k-1}\ |\ x_k,(y_0,\ldots,y_{k-1}))\ dx_{k-1}$$

Our next objective is to provide an analytic expression of these Markov transitions. To this end, we observe that

$$p(x_{k-1}\ |\ x_k,(y_0,\ldots,y_{k-1}))\ \propto\ p(x_k\ |\ x_{k-1})\ p(x_{k-1}\ |\ (y_0,\ldots,y_{k-1}))$$

and

$$\eta_{k-1}(dx_{k-1}) = p(x_{k-1}\ |\ (y_0,\ldots,y_{k-1}))\ dx_{k-1}$$

In terms of Boltzmann-Gibbs transformation, we have the Bayes' formula

$$\mathbb{M}_{k,\eta_{k-1}}(x_k,dx_{k-1}) = \Psi_{G_{x_k,k}}\left(\widehat{\eta}_k\right)(dx_{k-1}) \quad \text{with} \quad G_{x_k,k}(x_{k-1}) = p(x_k\ |\ x_{k-1})$$

with the optimal filter distribution $\widehat{\eta}_k$ introduced in (8.25). We recall that

$$p(x_k\ |\ x_{k-1})\ dx_k = \mathcal{N}\left(A_k x_{k-1} + a_k, R_k\right)(dx_k) = g_{(m_k^X(x_{k-1}),R_k)}(x_k)\ dx_k$$

 Using the Gaussian-updating formula (8.29) we have

$$\mathbb{M}_{k,\eta_{k-1}}(x_k,.) = \Psi_{g_{(x_k-a_k,R_k)}(A_k.)}\left(\mathcal{N}\left(\widehat{X}_{k-1},P_{k-1}\right)\right) = \mathcal{N}\left(\widehat{m}_{x_k,k}\left(\widehat{X}_{k-1},P_{k-1}\right),\widehat{\Sigma}_k(P_{k-1})\right)$$

with the parameters

$$\begin{aligned}\widehat{m}_{x_k,k}\left(x_{k-1},P_{k-1}\right) &= x_{k-1}+P_{k-1}A_k'(P_k^-)^{-1}\ \left(x_k-(A_kx_{k-1}+a_k)\right)\\ \widehat{\Sigma}_k(P_{k-1}) &= (I-P_{k-1}A_k'(P_k^-)^{-1}A_k)P_{k-1}\end{aligned}$$

 This shows that (8.32) is the distribution of the backward random trajectories

$$\widetilde{X}_n^{(n)} \to \widetilde{X}_{n-1}^{(n)} \to \ldots \to \widetilde{X}_1^{(n)} \to \widetilde{X}_0^{(n)}$$

defined by the equations

$$\left\{\begin{array}{rcl}\widetilde{X}_p^{(n)} &=& \widehat{m}_{\widetilde{X}_{p+1},p+1}\left(\widehat{X}_p\right)+\widetilde{W}_p\\ &=& \widehat{X}_p+P_pA_{p+1}'\left(P_{p+1}^-\right)^{-1}\ \left(\widetilde{X}_{p+1}^{(n)}-\left[A_{p+1}\widehat{X}_p+a_{p+1}\right]\right)+\widetilde{W}_p\\ \widetilde{X}_n^{(n)} &\sim& \mathcal{N}\left(\widehat{X}_n^-,P_n^-\right)\end{array}\right. \tag{8.33}$$

with a sequence $\widetilde{W}_p$ of i.i.d. centered Gaussian variables with covariance matrices

$$\Sigma_p := \left(I-P_pA_{p+1}'\left(P_{p+1}^-\right)^{-1}A_{p+1}\right)P_p$$

The conditional mean and covariance matrices of this Gaussian linear model

$$\begin{aligned}\overline{X}_p^{(n)} &= \mathbb{E}\left(\widetilde{X}_p^{(n)}\mid(Y_0,\ldots,Y_{n-1})\right)\\ \Sigma_p^{(n)} &= \mathbb{E}\left(\left(\widetilde{X}_p^{(n)}-\overline{X}_p^{(n)}\right)\left(\widetilde{X}_p^{(n)}-\overline{X}_p^{(n)}\right)'\mid(Y_0,\ldots,Y_{n-1})\right)\end{aligned}$$

satisfy the backward recursive formula

$$\left\{\begin{array}{rcl}\overline{X}_p^{(n)} &=& \widehat{m}_{\overline{X}_{p+1},p+1}\left(\widehat{X}_p\right)\\ \Sigma_p^{(n)} &=& P_p+P_pA_{p+1}'\left(P_{p+1}^-\right)^{-1}\ \left(\Sigma_{p+1}^{(n)}-P_{p+1}^-\right)\left(P_{p+1}^-\right)^{-1}A_{p+1}P_p\end{array}\right.$$

with final time horizon condition $\left(\overline{X}_n^{(n)},\Sigma_n^{(n)}\right)=\left(\widehat{X}_n^-,P_n^-\right)$.

8.4 Hidden Markov chain models

8.4.1 Boltzmann posterior and Metropolis-Hastings models

Suppose we are given a signal-observation model (8.24) with kinetic parameters

$$(a_n,A_n,B_n,c_n,C_n,D_n)=(a_n(\Theta),A_n(\Theta),B_n(\Theta),c_n(\Theta),C_n(\Theta),D_n(\Theta))$$

that depend on some random variable Θ with a given distribution ν on some state space Ξ.

The posterior distributions of the signal trajectories given the observation sequence and the parameter Θ can be computed using the forward Kalman filter and the backward smoother presented in section 8.3.1 and in section 8.3.2.

 In addition, using (8.30) we have

$$\begin{aligned}\mathbb{P}\left((Y_0,\ldots,Y_n)\in d(y_0,\ldots,y_n)\mid\Theta=\theta\right) &= p_n(y_0,\ldots,y_n\mid\theta)\;dy_0\ldots dy_n\\ &= \left\{\prod_{k=0}^{n} g_{(C_k(\theta)\widehat{X}_{\theta,k}^{-}+c_k(\theta),\Sigma_k(P_{\theta,k}^{-}))}(y_k)\right\}\;dy_0\ldots dy_n\end{aligned}$$

where $(\widehat{X}_{\theta,k}^{-},P_{\theta,k}^{-})$ stands for the conditional expectation and the conditional covariance matrix

$$\mathrm{Law}(X_k\mid(Y_0,\ldots,Y_{k-1})=(y_0,\ldots,y_{k-1}),\;\Theta=\theta) \;=\; \mathcal{N}\left(\widehat{X}_{\theta,k}^{-},P_{\theta,k}^{-}\right)$$

computed using the Kalman recursions.

Using Bayes' rule, we prove that the posterior distribution of Θ given a fixed sequence of observations $\boldsymbol{Y_n}=(Y_0,\ldots,Y_n)=(y_0,\ldots,y_n)=\boldsymbol{y_n}$ is given by the Boltzmann-Gibbs measure

$$\mathbb{P}\left(\Theta\in d\theta\mid\boldsymbol{Y_n}=\boldsymbol{y_n}\right)\propto\left\{\prod_{k=0}^{n}h_k(\theta)\right\}\;\nu(d\theta)$$

with the likelihood potential functions

$$h_k(\theta):=g_{(C_k(\theta)\widehat{X}_{\theta,k}^{-}+c_k(\theta),\Sigma_k(P_{\theta,k}^{-}))}(y_k)$$

The sampling of these target measures can be performed using the Metropolis-Hasting algorithms developed in section 7.2. An alternative and more powerful particle simulation technique based on Metropolis-Hasting sampling schemes equipped with recycling mechanisms is provided in section 9.1.5.

8.4.2 Conjugate priors and Gibbs samplers

We start with a simple calibration model in which the parameters a_n in the linear Gaussian model (8.24) is a fixed and unknown parameters in $\mathbb{R}^p$. To simplify the presentation, we further assume that $(B_n,D_n)=(Id,Id)$ and (W_n,V_n) are centered independent Gaussian random variables with time homogeneous covariance matrices $(R_n^v,R_n^w)=(Q,R)$. In this situation, given some random variable Θ with distribution ν on $\mathbb{R}^d$ the signal model (8.24) is now given by

$$X_n=A_n\;X_{n-1}+\Theta+W_n \tag{8.34}$$

We fix a final time horizon n and a random path $X := (X_0, \ldots, X_n) = x$. Using Bayes' rule, the conditional distribution of Θ given is given by the formula

$$\mathbb{P}(\Theta \in d\theta \mid X = x) \propto \left\{ \prod_{1 \leq k \leq n} h_k(\theta) \right\} \nu(d\theta) \propto (\Psi_{h_n} \circ \ldots \circ \Psi_{h_1})(\nu)(d\theta)$$

with the collection of likelihood functions h_k defined by

$$h_k(\theta) := g_{(x_k - A_k x_{k-1}, R)}(\theta) \left(\overset{cf.(8.12)}{=} G_{\Delta x_k}(\theta) \quad \text{with} \quad \Delta x_k = x_k - A_k x_{k-1} \quad \text{and} \quad R_0 = R \right)$$

In the above displayed formula, we recall that $g_{(m,R)}$ are the Gaussian density functions defined in (8.11).

We further assume that the prior distribution $\nu = \mathcal{N}(m_1, P_1)$ this model has the same form as the one discussed in (8.14). We let π be the conditional distribution

$$\pi = \text{Law}((\Theta, X) \mid Y = y)$$

By construction, we have

$$\pi(d(\theta, x)) = \pi_1(d\theta)\ L_{1,2}(\theta, dx) \ = \ \pi_2(dx)\ L_{1,2}(x, d\theta)$$

with the marginal measures $(\pi_1, \pi_2) = (\text{Law}(\Theta | Y = y), \text{Law}(X | Y = y))$, and the conditional distributions

$$\begin{aligned} L_{1,2}(\theta, dx) &= \mathbb{P}(X \in dx \mid \Theta = \theta,\ Y = y) \\ L_{2,1}(x, d\theta) &= \mathbb{P}(\Theta \in d\theta \mid X = x,\ Y = y) = \mathbb{P}(\Theta \in d\theta \mid X = x) \end{aligned}$$

The Gibbs sampler (8.3) associated with these disintegration models is based on sampling sequentially the conditional distributions $(L_{1,2}, L_{2,1})$:

- The sampling of $L_{1,2}$ can be performed using the backward Kalman smoothers developed in section 8.3.2.
- The transition $L_{2,1}$ can be sampled using the conjugate mappings presented in (8.14).

The same analysis can be developed when the parameters a_n are known and the covariance matrix parameter $\Theta = \theta$ of the signal Gaussian perturbation $W_n = W_n^\theta \sim \mathcal{N}(0, \theta)$ are unknown. In this situation, given Θ, the signal model (8.24) is given by

$$X_n = A_n\ X_{n-1} + a_n + W_n^\Theta$$

Arguing as above, we have

$$\mathbb{P}(\Theta \in d\theta \mid X = x) \propto \left\{ \prod_{1 \leq k \leq n} G_{\Delta_k(x)}(\theta) \right\} \nu(d\theta) \propto \left(\Psi_{G_{\Delta_n(x)}} \circ \ldots \circ \Psi_{G_{\Delta_1(x)}} \right)(\nu)(d\theta)$$

with the collection of likelihood functions

$$g_{(A_k x_{k-1}+a_k,\theta)}(x_k) := g_{(0,\theta)}(\Delta_k(x)) \overset{cf.(8.12)}{=} G_{\Delta_k(x)}(\theta)$$

with

$$\Delta_k(x) = x_k - A_k x_{k-1} - a_k \quad \text{and} \quad R_0 = \theta$$

Assuming that the prior distribution $\nu = \mathcal{IW}_d(v, W)$ this model has the same form as the one discussed in (8.16) and (8.18).

In more general situations, the pair signal-observation model $\mathcal{X}_n = (X_n, Y_n)$ depend on some unknown parameter. These hidden Markov chain problems are better described in an abstract framework. We fix the time horizon $n \in \mathbb{N}$, and we let $E = (E_0 \times \ldots \times E_n)$ be the state space of the random trajectories $\mathcal{X} = (\mathcal{X}_0, \ldots, \mathcal{X}_n)$ of some conditional Markov chain $\mathcal{X}_k$ evolving in some state spaces E_k w.r.t. some random parameter Θ with some distribution $\pi_1(d\theta)$ on some state space Ξ.

We further assume that given $\Theta = \theta$ the elementary Markov transitions $M_{\theta,k}(x_{k-1}, dx_k)$, and the initial distribution $\eta_{\theta,0}(dx_0)$, have some density $m_{\theta,k}(x_{k-1}, x_k)$, and $m_{\theta,0}(x_0)$, w.r.t. some reference distributions $\lambda_k(dx_k)$ on the state spaces E_k. We consider the likelihood functions

$$G_{0,x_0}(\theta) = m_{\theta,0}(x_0) \quad \text{and} \quad \forall 1 \leq k \leq n \quad G_{k,x}(\theta) = m_{\theta,k}(x_{k-1}, x_k)$$

For instance, let us suppose that the parameters $(a_n, c_n) = (a, c) = \theta = (\theta^{(1)}, \theta^{(2)}) \in \mathbb{R}^{p+q}$ in the signal observation model $\mathcal{X}_n = (X_n, Y_n)$ discussed above are unknown. In this situation, given $\Theta = (\Theta^{(1)}, \Theta^{(2)})$ the Markov transition of the chain $\mathcal{X}$ are given by the equations

$$\left\{ \begin{array}{rcl} X_n & = & A_n\ X_{n-1} + \Theta^{(1)} + W_n \\ Y_n & = & C_n\ X_n \quad + \Theta^{(2)} + V_n \end{array} \right.$$

To describe with some precision the elementary transitions of this chain, we observe that

$$g_{(r,R)}(\theta^{(1)})\ g_{(q,Q)}(\theta^{(2)}) = g_{(s,S)}\,(\theta)$$

with the parameters

$$s = \left(\begin{array}{c} r \\ q \end{array} \right) \qquad \theta = \left(\begin{array}{c} \theta^{(1)} \\ \theta^{(2)} \end{array} \right) \quad \text{and} \quad S := \left(\begin{array}{cc} R & 0 \\ 0 & Q \end{array} \right)$$

In this notation, we have

$$\mathbb{P}\left((X_n, Y_n) \in d(x_n, y_n) \mid (X_{n-1}, Y_{n-1})\ , \Theta\right)$$

$$= \underbrace{g_{(A_n X_{n-1}+\Theta^{(1)},R)}(x_n) \times g_{(y_n-\Theta^{(2)},Q)}(C_n x_n)}_{m_{\Theta,n}((X_{n-1},Y_{n-1}),(x_n,y_n))}\ dx_n\ dy_n$$

Recalling that with

$$g_{(y_k-\theta^{(2)},Q)}(C_k x_k) = g_{(y_k-C_k x_k,Q)}(\theta^{(2)}) \quad \text{and} \quad g_{(A_k x_{k-1}+\theta^{(1)},R)}(x_k) = g_{(x_k-A_k x_{k-1},R)}(\theta^{(1)})$$

we find that

$$G_{k,(x,y)}(\theta) = g_{(\Delta_k(x,y),S)}(\theta) \quad \text{with} \quad \Delta_k(x,y) = \left(\begin{array}{c} x_k - A_k x_{k-1} \\ y_k - C_k x_k \end{array} \right)$$

To get one step further in our discussion, we fix the time parameter n, and we let π be a target distribution on $S = (\Theta \times E)$ of the form

$$\pi(d(\theta, x)) = \pi_1(d\theta)\ L_{1,2}(\theta, dx) = \pi_2(dx)\ L_{2,1}(x, d\theta)$$

with

$$L_{2,1}(x,d\theta) \propto \left[\prod_{0\leq k\leq n} G_{k,x}(\theta)\right]\ \pi_1(d\theta)\ \propto\ \Psi_{G_{n,x}} \circ \Psi_{G_{n-1,x}} \circ \ldots \circ \Psi_{G_{0,x}}(\pi_1)(d\theta)$$

We further assume that $\pi_1 = \nu_h \in \mathcal{P}$ belongs to some class of prior distributions $\mathcal{P}$ indexed by some (prior) hyper-parameter set $\mathcal{H} \ni h$, which are conjugate to some classes of likelihood functions $\mathcal{G}_k \ni G_k\ :\ (x,\theta) \mapsto G_{k,x}(\theta)$. In this case, there exists some conjugacy mapping

$$H_k\ :\ (h,x) \in (\mathcal{H} \times E) \mapsto H_{k,x}(h) \in \mathcal{H}$$

such that

$$\Psi_{G_{0,x}}(\nu_h) = \nu_{H_{0,x}(h)}\ \Rightarrow\ \Psi_{G_{1,x}}\left(\Psi_{G_{0,x}}(\nu_h)\right) = \nu_{H_{1,x}(H_{0,x}(h))} \Rightarrow L_{2,1}(x,d\theta) = \nu_{H_{0,n,x}(h)}$$

with the composition semigroup

$$H_{0,n,x} = H_{n,x} \circ \ldots \circ H_{0,x}$$

8.5 Computational physics

8.5.1 Molecular dynamics

Molecular dynamics simulation is concerned with the analysis of the fluctuations, and the conformal changes of proteins and nucleic acids in biological molecules. The central problem is to understand the macroscopic properties of a molecule through the simulation of a microscopic system of atomic interacting particles in a given force field model. More formally, we consider the microscopic evolution of a many-body system formed by k atomic particles in the Euclidian space $E = \mathbb{R}^3$ with possibly k different masses $m = (m_i)_{1\leq i\leq k}$. Their spatial positions, and their velocities are denoted by the letters $q = (q_i)_{1\leq i\leq k}$, and $p = (p_i)_{1\leq i\leq k}$. These particles moves under the influence of some external forces $F_i(q)$ according to the Newton's second law

$$m_i\ \frac{d^2 q_i}{dt^2} = F_i(q) \tag{8.35}$$

The velocity vector

$$p_i = \begin{pmatrix} p_i^1 \\ p_i^2 \\ p_i^3 \end{pmatrix} = m_i\ \frac{dq_i}{dt} = m_i \begin{pmatrix} \dfrac{dq_i^1}{dt} \\ \dfrac{dq_i^2}{dt} \\ \dfrac{dq_i^3}{dt} \end{pmatrix}$$

is called the particle momenta of the system, and the couple $x = (q,p)$ is called the phase vector.

We further assume that the force field is conservative, in the sense that

$$F(q) = -\nabla_q V(q) = \left(-\frac{\partial V}{\partial q_i}(q)\right)_{1\leq i\leq k}$$

for some interparticle potential function $V : E^k \to \mathbb{R}$.

 In this situation, we can reformulate the evolution equations (8.35) in terms of the Hamiltonian or energy functional

$$H(q,p) = \sum_{i=1}^{k} \frac{\|p_i\|^2}{2m_i} + V(q_1,\ldots,q_k) \tag{8.36}$$

with the following equations

$$\left\{ \begin{array}{rcl} \dfrac{dq_i}{dt} & = & \dfrac{p_i}{m_i} = \dfrac{\partial H}{\partial p_i}(q,p) \\ \dfrac{dp_i}{dt} & = & F_i(q) = -\dfrac{\partial V}{\partial q_i}(q) = -\dfrac{\partial H}{\partial q_i}(q,p) \end{array} \right. \tag{8.37}$$

We notice that these evolution equations are time reversible, in the sense that they have the same form if we consider the time transformation $\tau(t) = -t$. In other words, the microscopic physics doesn't depends on the time flow direction. We also notice the conservation property

$$\frac{d}{dt}H(q,p) = \sum_{i=1}^{k} \left[\frac{\partial H}{\partial q_i}(q,p)\frac{dq_i}{dt} + \frac{\partial H}{\partial p_i}(q,p)\frac{dp_i}{dt} \right] = 0 \tag{8.38}$$

In the above display we have used the conventions

$$\frac{\partial H}{\partial q_i} = \left(\frac{\partial H}{\partial q_i^1}, \frac{\partial H}{\partial q_i^2}, \frac{\partial H}{\partial q_i^3} \right) \quad \text{and} \quad \frac{dq_i}{dt} = \left(\begin{array}{c} \dfrac{dq_i^1}{dt} \\ \dfrac{dq_i^2}{dt} \\ \dfrac{dq_i^3}{dt} \end{array} \right) \tag{8.39}$$

and

$$\frac{\partial H}{\partial p_i} = \left(\frac{\partial H}{\partial p_i^1}, \frac{\partial H}{\partial p_i^2}, \frac{\partial H}{\partial p_i^3} \right) \quad \text{and} \quad \frac{dp_i}{dt} = \left(\begin{array}{c} \dfrac{dp_i^1}{dt} \\ \dfrac{dp_i^2}{dt} \\ \dfrac{dp_i^3}{dt} \end{array} \right) \tag{8.40}$$

We also mention that for $k = 1$ and $V(q) = \frac{k}{2}\, q^2$, for some $k \geq 0$, the system (8.37) reduces to the linearized pendulum

$$\left. \begin{array}{rcl} \dfrac{dq}{dt} & =: & q' = \dfrac{p}{m} \\ \dfrac{dp}{dt} & =: & p' = -kq \end{array} \right\} \Rightarrow \frac{d^2q}{dt^2} + \omega^2 q = 0 \quad \text{with} \quad \omega = \sqrt{\frac{k}{m}}$$

The solution of this system takes the form

$$q(t) = q(0)\cos(\omega t) + \frac{q'(0)}{\omega}\ \sin(\omega t)$$

Solid and liquid states of rare-gas elements with closed shell configurations only involves particle interacting with weak van de Waals bonds in terms of the pair-potential function

$$V(q_1,\ldots,q_k) = \sum_{1\leq i<j\leq k} V_{LJ}(\|q_j - q_i\|)$$

with the Lennard Jones potential functions

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\tau}{r}\right)^{12} - \left(\frac{\tau}{r}\right)^{6} \right] \tag{8.41}$$

The parameter ϵ represents the depth of the potential well, and τ the finite distance at which the interaction potential becomes null. Notice that

$$\inf_r V_{LJ}(r) = V_{LJ}(2^{1/6}\tau) = -\epsilon$$

This shows that for $r \geq 2^{1/6}\tau$ the potential is attractive, and repulsive for $r \leq 2^{1/6}\tau$.

The term $(\tau/r)^{12}$ describes the short range Pauli repulsion forces due to overlapping electron orbitals, while the term $(\tau/r)^6$ represents the attraction and the van der Waals dispersion forces at long range distances.

The repulsion term has no real theoretical foundations, it is sometimes replaced by the Buckingham exponential-6 potential $\exp(-r/\tau)$. To avoid the degeneracy of the Lennard Jones potential at short range distances, we often use cut-off techniques. For instance, we can replace $V_{LJ}(r)$ by

$$\overline{V}_{LJ}(r) = (V_{LJ}(r) - V_{LJ}(r_c))\ 1_{r<r_c}$$

or by

$$\overline{V}_{LJ}(r) = \big(V_{LJ}(r) - V_{LJ}(r_c) - V'_{LJ}(r_c)(r - r_c)\big)\ 1_{r<r_c}$$

for some well chosen cut-off radius r_c. For instance, the Wayne-Chandler-Anderson potential is given by

$$r_c = 2^{1/6}\tau \implies \overline{V}_{WCA}(r) = \overline{V}_{LJ}(r) = (V_{LJ}(r) + \epsilon)\ 1_{r<2^{1/6}\tau}$$

We associate with the Hamiltonian function (8.36), the canonical measures on the phase space

$$\mu_\beta(dx) = \frac{1}{\mathcal{Z}_\beta}\ e^{-\beta H(x)}\ dx \tag{8.42}$$

where $\mathcal{Z}_\beta$ is a normalizing constant, and $dx = d(q,p) = dq \times dp$ stands for the Lebesgue measure on $\mathbb{R}^{3k+3k}$, and $x = (q,p)$ stands for a given point in the phase space.

We also consider the q-marginal measures

$$\overline{\mu}_\beta(dq) = \frac{1}{\overline{\mathcal{Z}}_\beta}\ e^{-\beta V(q)}\ dq \tag{8.43}$$

where $\mathcal{Z}_\beta$ is a normalizing constant, and dq stands for the the Lebesgue measure on the position space $\mathbb{R}^{3k}$.

In this notation, the measure μ_β is given by the product formula

$$\mu_\beta(d(q,p)) \quad = \quad \left[\prod_{1\leq i\leq k} \frac{1}{\sqrt{2\pi m_i/\beta}}\ e^{-\beta\frac{p_i^2}{2m_i}}\ dp_i \right]\ \overline{\mu}_\beta(dq)$$

The Boltzmann-Gibbs measures μ_β, and respectively $\overline{\mu}_\beta$, can be interpreted as the invariant measure of the Langevin type stochastic dynamics

$$\left\{\begin{array}{rcl} dq_i &=& \beta \overbrace{\frac{\partial H}{\partial p_i}(q,p)}^{p_i/m_i}\ dt \\ dp_i &=& -\beta \underbrace{\left[\frac{\partial H}{\partial q_i}(q,p)+\sigma^2\frac{\partial H}{\partial p_i}(q,p)\right]}_{=\frac{\partial V}{\partial q_i}(q)+\sigma^2\ p_i/m_i} dt+\sigma\sqrt{2}\ dW_t^i \end{array}\right. \tag{8.44}$$

and respectively

$$dq_i = -\beta\frac{\partial V}{\partial q_i}(q)\ dt+\sqrt{2}\ dW_t^i \tag{8.45}$$

where $(W_t^i)_{1\leq i\leq k}$ stands for k independent Brownian motions $W_t^i=\left(W_t^{i,j}\right)_{1\leq j\leq 3}$ on $\mathbb{R}^3$.

The additional external Brownian forces represent the fluctuations of the many-body system, balanced by dissipative and viscous damping forces. In both cases, one can show that the Markov evolution semigroups of these diffusions have a density w.r.t. the Lebesgue measure on $\mathbb{R}^{3k}$ or on $\mathbb{R}^{3k+3k}$. The fact that these density are smooth relies on more sophisticated stochastic analysis tools, including Malliavin Calculus and differential geometry [150, 330, 372].

We check that μ_β, and $\overline{\mu}_\beta$ are the invariant measures of these diffusion models using the infinitesimal generators of the diffusion processes (8.44) and (8.45), given respectively on the set of smooth function f on $\mathbb{R}^{3k+3k}$ by the formulae

$$L_\beta(f)=\beta\sum_{i=1}^k\left[\frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q_i}-\left(\frac{\partial H}{\partial q_i}+\sigma^2\frac{\partial H}{\partial p_i}\right)\frac{\partial f}{\partial p_i}\right]+\sigma^2\sum_{i=1}^k\frac{\partial^2 f}{\partial p_i^2}$$

and for any smooth function g on $\mathbb{R}^{3k}$ by

$$\overline{L}_\beta(g)=-\beta\sum_{i=1}^k\frac{\partial V}{\partial q_i}\frac{\partial g}{\partial q_i}+\sum_{i=1}^k\frac{\partial^2 g}{\partial q_i^2}=e^{\beta V}\ \sum_{i=1}^k\frac{\partial}{\partial q_i}\left(e^{-\beta V}\ \frac{\partial g}{\partial q_i}\right)$$

In the above display we slightly abuse the notation dropping the transposition operator $(.)'$ in the differential of the functions f and g. For instance, using the conventions (8.39) and (8.40), we have that

$$\frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q_i}:=\frac{\partial H}{\partial p_i}\left(\frac{\partial f}{\partial q_i}\right)'=\frac{\partial H}{\partial p_i^1}\frac{\partial f}{\partial q_i^1}+\frac{\partial H}{\partial p_i^2}\frac{\partial f}{\partial q_i^2}+\frac{\partial H}{\partial p_i^3}\frac{\partial f}{\partial q_i^3}$$

To simplify the presentation, we have also denoted by $\frac{\partial^2}{\partial p_i^2}$ the Laplacian operator on $\mathbb{R}^3$; that is, we have that

$$\frac{\partial^2 f}{\partial p_i^2}=\frac{\partial^2 f}{\partial (p_i^1)^2}+\frac{\partial^2 f}{\partial (p_i^2)^2}+\frac{\partial^2 f}{\partial (p_i^3)^2}$$

and

$$\frac{\partial}{\partial q_i}\left(e^{-\beta V}\ \frac{\partial g}{\partial q_i}\right):=\frac{\partial}{\partial q_i^1}\left(e^{-\beta V}\ \frac{\partial g}{\partial q_i^1}\right)+\frac{\partial}{\partial q_i^2}\left(e^{-\beta V}\ \frac{\partial g}{\partial q_i^2}\right)+\frac{\partial}{\partial q_i^3}\left(e^{-\beta V}\ \frac{\partial g}{\partial q_i^3}\right)$$

In this stochastic framework, the conservation properties (8.38) takes the following form.

 For any $\beta \in \mathbb{R}$, we have

$$\mu_\beta L_\beta = 0 \quad \text{and} \quad \overline{\mu}_\beta \overline{L}_\beta = 0$$

In addition $\overline{\mu}_\beta$ is $\overline{L}_\beta$-reversible, in the sense that for any smooth couple of functions (g, h) with compact support on $\mathbb{R}^{3k}$ we have

$$\overline{\mu}_\beta \left(g \ \overline{L}_\beta(h) \right)) = \overline{\mu}_\beta \left(\overline{L}_\beta(g) \ h \right)$$

Proof :
By a simple integration by part formula, for any smooth function f with compact support on $\mathbb{R}^{3k+3k}$ we check that

$$\begin{aligned}
&\int e^{-\beta H(x)} \ L_\beta(f)(x) \ dx \\
&= -\beta \sum_{i=1}^k \ \int f(x) \ \frac{\partial}{\partial q_i} \left(e^{-\beta H} \ \frac{\partial H}{\partial p_i} \right)(x) \ dx \\
&\qquad +\beta \sum_{i=1}^k \ \int f(x) \ \frac{\partial}{\partial p_i} \left(e^{-\beta H} \ \left(\frac{\partial H}{\partial q_i} + \sigma^2 \frac{\partial H}{\partial p_i} \right) \right)(x) \ dx \\
&\qquad\qquad +\sigma^2 \sum_{i=1}^k \int f(x) \frac{\partial^2}{\partial p_i^2} \left(e^{-\beta H} \right)(x) \ dx
\end{aligned}$$

This implies that

$$\begin{aligned}
&\mu_\beta \left(L_\beta(f) \right) \\
&= \sum_{i=1}^k \ \mu_\beta \left\{ f \ \left[\left(\beta^2 \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} - \beta \frac{\partial^2 H}{\partial q_i \partial p_i} \right) - \beta^2 \ \frac{\partial H}{\partial p_i} \left(\frac{\partial H}{\partial q_i} + \sigma^2 \frac{\partial H}{\partial p_i} \right) \right] \right\} \\
&+ \sum_{i=1}^k \ \mu_\beta \left\{ f \ \left[\beta \left(\frac{\partial^2 H}{\partial q_i \partial p_i} + \sigma^2 \frac{\partial^2 H}{\partial p_i^2} \right) - \sigma^2 \beta \frac{\partial^2 H}{\partial p_i^2} + \sigma^2 \beta^2 \left(\frac{\partial H}{\partial p_i} \right)^2 \right] \right\} = 0
\end{aligned}$$

In much the same way, for any smooth functions (g, h) with compact support on $\mathbb{R}^{3k}$ we find that

$$\begin{aligned}
\int e^{-\beta V(q)} g(q) \ \overline{L}_\beta(h)(q) \ dq &= \sum_{i=1}^k \int g(q) \ \frac{\partial}{\partial q_i} \left(e^{-\beta V} \ \frac{\partial h}{\partial q_i} \right)(q) \ dq \\
&= -\sum_{i=1}^k \int e^{-\beta V(q)} \ \frac{\partial g}{\partial q_i}(q) \frac{\partial h}{\partial q_i}(q) \ dq \\
&= \sum_{i=1}^k \int h(q) \ \frac{\partial}{\partial q_i} \left(e^{-\beta V} \ \frac{\partial g}{\partial q_i} \right)(q) \ dq
\end{aligned}$$

This clearly ends the proof of the lemma. ∎

The stability properties of the Langevin models (8.44) and (8.45) can be analyzed using the tools developped in section 5.7, section 5.8. More precisely, we first check that the semigroup P_t of these diffusion models have a smooth density w.r.t. the Lebesgue measure. This will ensure that P_t satisfies the Dobrushin local contraction condition (4.39) for any $t > 0$. The second step is to find a judicious

Lyapunov function satisfying the condition (5.45). By (5.46), these two properties ensure that the law of the random states of the Langevin models (8.44) and (8.45) converge exponentially fast, as the time parameter tends to infinity, to the invariant measures (8.42) and (8.43). Nevertheless, up to our knowledge most of the Lyapunov functions developed in the literature on Langevin diffusions require that the potential functions behave as polynomials at infinity. These techniques cannot be used to analyze the Lennard-Jones potential functions presented in (8.41). The only work in this direction seems to be the article by B. Cooke, J.C. Mattingly, S.A. McKinley, and S.C. Schmidler [150] on a reduced two-dimensional Langevin diffusion model.

8.5.2 The Schrödinger equation

A physical derivation

The Schrödinger equation is the quantum mechanics version of the Newton's second law of motion of classical mechanics (the mass times the acceleration is the sum of the forces). This equation represent the wave function (a.k.a. the quantum state) evolution of some physical system, including molecular, atomic of subatomic systems, as well as macroscopic systems like the universe [532].

In 1924 de Broglie made the hypothesis that if light waves of frequency ω behave as a population of particles of energy $E = \hbar\omega$, then massive particles with energy E can also behave like waves of frequency $\omega = E/\hbar$. More precisely, the wave function of a free particle of momentum $p = \hbar k$ and energy

$$E = \frac{p^2}{2m} = \hbar\omega \ \Rightarrow E = \frac{k^2\hbar^2}{2m} = \frac{p^2}{2m}$$

has the following form

$$\psi(t,x) = \psi_0 \ e^{i(kx-\omega t)}$$

This wave function is the result of two traveling waves in the x and t directions.

An elementary computation shows that

$$\frac{\partial\psi}{\partial x} = ik \ \psi \Rightarrow -\frac{\hbar^2}{2m} \ \frac{\partial^2\psi}{\partial x} = k^2 \ \frac{\hbar^2}{2m} \ \psi = \frac{p^2}{2m} \ \psi = E \ \psi$$

and

$$i\hbar\frac{\partial\psi}{\partial t} = \hbar\omega \ \psi = E \ \psi$$

from which we conclude that

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x} = E \ \psi = i\hbar\frac{\partial\psi}{\partial t}$$

and

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x}$$

Extending these wave functions to particles motions in a potential energy $V(x)$ the energy E is the sum of the kinetic and the potential energies

$$E = \frac{p^2}{2m} + V(x)$$

Assuming that the above equations are valid in this case, we obtain the time dependent Schrödinger wave equation

$$i\hbar\frac{\partial\psi}{\partial t} = E \ \psi = -\frac{p^2}{2m}\psi + V\psi = -\frac{\hbar^2}{2m} \ \frac{\partial^2\psi}{\partial x} + \ V\psi$$

Rewritten in a slightly different form, the Schrödinger wave equation equation takes the following form

$$i\hbar\frac{\partial\psi}{\partial t} = -L^V(\psi) \quad \text{with the Schrödinger operator} \quad L^V = \frac{\hbar^2}{2m} \ \frac{\partial^2}{\partial x} - V$$

A formal change of time coordinate $t = -i\tau\hbar$ and $u(\tau, x) = \psi(-i\tau\hbar, x)$ transforms the above equation into a the heat type equation

$$\frac{\partial u}{\partial \tau} = L^V(u) \tag{8.46}$$

In physics, this change of coordinate is sometimes called a Wick rotation of the time axis, and the resulting equation is often referred as the Schrödinger equation in imaginary time.

Notice that the formal change of time coordinate $t = -i\tau$ and

$$v(\tau, x) := \psi(-i\tau, x) = u(\tau/\hbar, x)$$

transforms the above equation into

$$\frac{\partial v}{\partial \tau}(\tau, x) = \frac{\hbar}{2m}\frac{\partial^2 v}{\partial x^2}(\tau, x) - \frac{1}{\hbar}V(x)v(\tau, x) \tag{8.47}$$

Feynman-Kac formulae

We consider a time homogeneous stochastic process X_τ on some state space S with infinitesimal generator L acting on some domain of functions $D(L)$.

We denote by Q_τ the integral operator defined for any bounded function f on S by the formula

$$Q_\tau(f)(x) := \mathbb{E}\left(f(X_\tau)\ \exp\left\{-\int_0^\tau V(X_s)ds\right\} \mid X_0 = x\right) \tag{8.48}$$

We also consider the Feynman-Kac measures γ_t and η_t defined by

$$\eta_t(f) = \gamma_t(f)/\gamma_t(1) \quad \text{and} \quad \gamma_t(f) = \mathbb{E}\left(f(X_\tau)\ \exp\left\{-\int_0^\tau V(X_s)ds\right\}\right) \tag{8.49}$$

In the further development of this section we implicitly assume that $Q_\tau(D(L))$ and $L(D(L))$ are subsets of $D(L)$. This condition depends on the regularity property of the generator L. For jump type infinitesimal generators this condition holds for any bounded potential function with $D(L) = \mathcal{B}(S)$. For diffusion type infinitesimal generators on $S = \mathbb{R}^d$, this condition holds for twice differentiable functions $D(L) = C_b^2(S)$ and bounded smooth potential functions.

We have the sg property

$$\forall s, t \geq 0 \qquad Q_{s+t} = Q_s Q_t \quad \text{and} \quad Q_0 = Id \tag{8.50}$$

In addition, the following evolution are satisfied for any $f \in D(L)$

$$\frac{\partial}{\partial \tau}Q_\tau(f) := Q_\tau(L^V(f)) = L^V(Q_\tau(f)) \tag{8.51}$$

 In particular, the function $u(\tau, x) := Q_\tau(f)(x)$ satisfies the equation

$$\frac{\partial u}{\partial \tau} = L^V(u) \quad \text{with} \quad L^V = L - V \tag{8.52}$$

These properties of Feynman-Kac semigroups have been proved in section 5.5. Next, we provide an alternative formal derivation of the evolution equations (8.51) and (8.52). For any $s \leq \tau$, using the Markov property, we prove that

$$Q_\tau(f)(x) = \mathbb{E}\left(\underbrace{\mathbb{E}\left(f(X_\tau)\ e^{-\int_s^\tau V(X_r)dr} \mid X_s \right)}_{=Q_{\tau-s}(f)(X_s)}\ e^{-\int_0^s V(X_r)dr} \mid X_0 = x \right)$$

This yields the sg property (8.50). Now we come to the proof of (8.51). We use the decomposition

$$\begin{aligned} & Q_{\tau+d\tau}(f)(x) - Q_\tau(f)(x) \\ &= \mathbb{E}\left(f(X_{\tau+d\tau})\ \left(e^{-\int_0^{\tau+d\tau} V(X_s)ds} - e^{-\int_0^\tau V(X_s)ds} \right) \mid X_0 = x \right) \\ &\qquad + \mathbb{E}\left((f(X_{\tau+d\tau}) - f(X_\tau))\ e^{-\int_0^\tau V(X_s)ds} \mid X_0 = x \right) \end{aligned}$$

The first term is given by

$$\begin{aligned} & \mathbb{E}\left(f(X_{\tau+d\tau})\ \left(e^{-\int_0^{\tau+d\tau} V(X_s)ds} - e^{-\int_0^\tau V(X_s)ds} \right) \mid X_0 = x \right) \\ &= \mathbb{E}\left(f(X_{\tau+d\tau})\ e^{-\int_0^\tau V(X_s)ds}\ \left(e^{-\int_\tau^{\tau+d\tau} V(X_s)ds} - 1 \right) \mid X_0 = x \right) \\ &\simeq \mathbb{E}\left((-V)(X_\tau)\ f(X_\tau)\ e^{-\int_0^\tau V(X_s)ds} \mid X_0 = x \right)\ d\tau = Q_\tau((-V)f)\ d\tau \end{aligned}$$

and the second one is given by

$$\begin{aligned} & \mathbb{E}\left((f(X_{\tau+d\tau}) - f(X_\tau))\ e^{-\int_0^\tau V(X_s)ds} \right) \\ &= \mathbb{E}\left((\mathbb{E}\left(f(X_{\tau+d\tau}) \mid X_\tau \right) - f(X_\tau))\ e^{-\int_0^\tau V(X_s)ds} \right) \\ &= \mathbb{E}\left(L(f)(X_\tau)\ e^{-\int_0^\tau V(X_s)ds} \right)\ d\tau = Q_\tau(L(f))\ d\tau \end{aligned}$$

This ends the proof of the first assertion. The r.h.s. of formula (8.51) comes from the fact that

$$Q_{\tau+d\tau} = Q_{d\tau}Q_\tau \implies Q_{\tau+d\tau} - Q_\tau = \underbrace{[Q_{d\tau} - Id]}_{\simeq\ L^V\ d\tau}\ Q_\tau$$

This ends the proof of the desired evolution equations.

The integral operators Q_τ can be made more explicit using the following formulae

$$\begin{aligned} & \mathbb{E}\left(f(X_\tau)\ e^{-\int_0^\tau V(X_s)ds} \mid X_0 = x \right) \\ &= \mathbb{E}\left(\mathbb{E}\left(e^{-\int_0^\tau V(X_s)ds} \mid X_0, X_\tau \right)\ f(X_\tau) \mid X_0 = x \right) \\ &= \int \underbrace{\mathbb{E}\left(e^{-\int_0^\tau V(X_s)ds} \mid X_0 = x,\ X_\tau = y \right)\ \mathbb{P}\left(X_\tau \in dy \mid X_0 = x \right)}_{:=Q_\tau(x,dy)}\ f(y) \end{aligned}$$

Using the sg property (8.50), the function $u(t,x) := Q_t(f)(x)$ satisfies the transport equation

$$u(s+t,x) := \int Q_s(x,dy)\ u(t,y)$$

In physics, the Feynman-Kac sg is also called the Green function, or the Feynman-Kac propagator.

Notice that

$$(8.51) \Rightarrow \frac{\partial^2}{\partial t^2} Q_t(f) = L^V \left(\frac{\partial}{\partial t} Q_t(f) \right) = (L^V)^2 (Q_t(f)) \Rightarrow \forall n \geq 1 \quad \frac{\partial^n}{\partial t^n} Q_\tau(f) \stackrel{t=0}{=} (L^V)^n f$$

Thus, formally we can write the evolution semigroup

$$Q_t = e^{-t\mathcal{H}} \quad \text{with the Hamiltonian operator} \quad \mathcal{H} = -L^V = -L + V \tag{8.53}$$

in the sense that

$$Q_t(f) = e^{tL^V} f = \sum_{n \geq 0} \frac{t^n}{n!} (L^V)^n (f) = e^{-t\mathcal{H}} f$$

8.5.3 The Ising model

The `Ising model` currently used in electromagnetism, statistical mechanics, as well as image processing is associated with the state space

$$S = \{-1, +1\}^E \qquad E = \{1, \ldots, L\} \times \{1, \ldots, L\}.$$

equipped with the uniform measure $\lambda(x) = 2^{-L^2}$. The lattice E is equipped with the following graph structure

$$j^1 = (i_1, i_2 + 1)\,, \qquad j^2 = (i_1 + 1, i_2)\,, \qquad j^3 = (i_1, i_2 - 1)\,, \qquad j^4 = (i_1 - 1, i_2)$$

around some state $(i_1, i_2) \in E$. Two neighbors $i, j \in E$ are denoted by $i \sim i'$.

The energy of a configuration $x \in S$ is given by the Hamiltonian function

$$V(x) = h \sum_{i \in E} x(i) - J \sum_{i \sim j} x(i)x(j), \tag{8.54}$$

The parameter $h \in \mathbb{R}$ represents the strength of an external magnetic field, and $J \in \mathbb{R}$ reflects the interaction degree between the sites.

In the Sherrington-Kirkpatrick model introduced in 1975 in their seminal article [535], the potential function is given by

$$V(\theta, x) := \sum_{1 \leq i \leq j \leq d} \theta_{i,j}\ x(i)\ x(j) + h \sum_{i=1}^{d} x(i)$$

where $\Theta_{i,j}$ are assumed to be i.i.d. centered gaussian random variables. More general disordered models can be defined in terms of random mappings $\Theta\ :\ (i,j) \in \{1, \ldots, d\}^2 \mapsto \Theta_{i,j}$.

Another way to extend the Ising model is to replace the spin state $\{-1,+1\}$ by some finite set of colors $C := \{c_1, \ldots, c_q\}$, for some $q \geq 1$. In this situation, the energy of a configuration $x \in S := C^E$ is given by the Hamiltonian function

$$V(x) = h \sum_{i \in E} x(i) - J \sum_{i \sim j} 1_{x(i)=x(j)}, \tag{8.55}$$

This model was introduced by Renfrey Potts in his 1951 PhD thesis [498]. A more physical model is given by the planar Potts model, also called the clock model suggested to him by his advisor Cyril Domb. In this model, the spins are replaced by angles

$$\forall 1 \leq k \leq q \qquad \theta_k = 2\pi q/k$$

and the energy of a configuration $x \in S := \{1, \ldots, q\}^E$ is given by

$$V(x) = -J \sum_{i \sim j} \cos\left(\theta_{x(i)} - \theta_{x(j)}\right),$$

The four-state clock model is sometimes termed the Ashkin-Teller model was introduced by in 1943 by Julius Ashkin and Edward Teller in [16]. The Boltzmann-Gibbs measures associated with these models are defined by

$$\pi(x) = \frac{1}{\mathcal{Z}_\beta} \, e^{-\beta V(x)} \, \lambda(x)$$

where β stands for some inverse temperature parameter, and $\mathcal{Z}_\beta$ some normalizing constant.

In the Gibbs model, for any fixed $i \in I$ we have

$$\begin{aligned} \pi(x) \quad \propto \quad & e^{-\beta h \; x(i) - \beta J \; x(i) \; \sum_{j \sim i} x(j)} \\ & \times \; e^{-\beta h \; \sum_{j \in I-i} x(j) - \beta J \; \sum_{j \sim k, \; j,k \in I-i} x(j) x(k)} \end{aligned}$$

This yields the disintegration formulae

$$\forall i \in I \qquad \pi(x) := \pi_i(x_i) \; L_{I-i,i}(x_{I-i}; x_i)$$

with the i-th marginal of π_i of π, and the conditional distribution

$$L_{I-i,i}(x_{I-i}; x_i) \propto \exp\left(-\beta h \; x_i - \beta J \; x_i \sum_{j \sim i} x_j \right)$$

More precisely, we have the following spin-site updates

$$\begin{aligned} L_{I-i,i}(x_{I-i}; \{1\}) &= 1 - L_{I-i,i}(x_{I-i}; \{-1\}) \\ &= \frac{e^{-\beta h - \beta J \; \sum_{j \sim i} x_j}}{e^{-\beta h - \beta J \; \sum_{j \sim i} x_j} + e^{+\beta h \; + \beta J \; \sum_{j \sim i} x_j}} \\ &= 1 / \left(1 + e^{2\beta \left[h + J \; \sum_{j \sim i} x_j \right]} \right) \end{aligned}$$

We associate with these models the Markov transitions K_i given for any $i \in I$ by

$$K_i(x, dy) := \delta_{x_{I-i}}(dy_{I-i}) \; L_{I-i,i}(y_{I-i}, dy_i) \tag{8.56}$$

The corresponding Gibbs sampler is the Markov chain with the elementary transition

$$M = \prod_{i \in I} K_i \qquad \text{in any order} \tag{8.57}$$

We can alternatively choose the Markov transitions

$$M = \frac{1}{\text{Card}(I)} \sum_{i \in I} K_i$$

Arguing as in section 7.3, one checks that M is reversible w.r.t. π, so that $\pi M = \pi$.

To sample the transition $x \rightsquigarrow y$ w.r.t. the Markov transition K_i given in (8.56), we sample a uniform r.v. U on $[0, 1]$ and we set

$$y = F^{(i)}(x)$$

with the random function $F^{(i)} \ : \ x \in S \to F^{(i)}(x) \in S$ defined by

$$\forall j \in I - i \qquad F^{(i)}(x)(j) = x(j)$$

and

$$F^{(i)}(x)(i) := 1_{[0, p_i(x)[}(U) - 1_{[p_i(x), 1]}(U) \tag{8.58}$$

with

$$p_i(x) := 1/\left(1 + e^{2\beta[h + J \ \sum_{j \sim i} x(j)]}\right) \tag{8.59}$$

When $J < 0$, the chance to pick the spin $+1$ increases as the number $j \sim i$ have the spin $+1$. This model is called an attractive spin system.

We equip the state space $S = \{-1, +1\}^I$, with $I = \{1, \ldots, L\}^2$ with the partial order

$$x \leq y \implies \forall i \in I \qquad x(i) \leq y(i)$$

The minimal and maximal states x_{min} and x_{max} are clearly given by

$$\forall i \in I \qquad x_{min}(i) = -1 \quad \text{and} \quad x_{max}(i) = +1$$

We also observe that

$$x \leq y \ \Rightarrow \ \forall i \in I \qquad \sum_{j \sim i} x(j) \leq \sum_{j \sim i} y(j)$$

Thus, when $J < 0$ the functions $F^{(i)}$ and $p_i(x)$ defined in (8.58) and (8.59) are such that

$$\begin{aligned} x \leq y \quad &\Rightarrow \quad \forall i \in I \qquad p_i(x) := 1/\left(1 + e^{2\beta[h - |J| \ \sum_{j \sim i} x(j)]}\right) \leq p_i(y) \\ &\Rightarrow \quad \forall i \in I \qquad F^{(i)}(x) \leq F^{(i)}(y) \end{aligned}$$

Given a sequence of independent functions $F^{(i)}$, with $i \in I$, the functions

$$F = \circ_{i \in I} F^{(i)} \quad \text{(in any order)} \quad \text{and} \quad F = \frac{1}{\text{Card}(I)} \sum_{i \in I} F^{(i)}$$

and monotone and compatible w.r.t. the Gibbs Markov transitions. These functions can be used to sample the Ising model using the Propp and Wilson technique.

8.5.4 Graph coloring models

We let $E = \{1, \dots, d\}$ can be the set of colors on the vertices of some graph $I = (\mathcal{V}, \mathcal{E})$. When $E = \{0, 1\}$ the color 0 can be interpreted as an empty site, and the color 1 as an occupied site. In this situation, a given configuration $x = (x(i))_{i \in I}$ can be interpreted as a collection of particles placed on the vertices $i \in I$ s.t. $x(i) = 1$.

We let $X = (X_i)_{i \in I}$ be some r.v. with distribution π on E^I. For any fixed $i \in I$, and any $x \in E^I$ we set

$$x_{I-i} = (x_j)_{j \in I - \{i\}} \quad \text{with} \quad I - i := \{j \in I \ : \ j \neq i\}$$

We assume that the following desintegration property is satisfied

$$\pi(dx) = \pi_{I-i}(dx_{I-i}) L_{I-i,i}(x_{I-i}, dx_i)$$

with the i-th marginals π_i of π and the conditional probability measure

$$L_{I-i,i}(x_{I-i}, dx_i) = \mathbb{P}\left(X_i \in dx_i \mid X_{I-i} = x_{I-i}\right)$$

In the above displayed formula, dx, resp. dx_i and dx_{I-i}, stand for an infinitesimal neighborhood of the point $x \in E^I$, resp. $x_i \in E$ and $x_{I-i} \in E^{I-i}$.

In the graph coloring model discussed above, we let A be the set of graph coloring such that two neighbor vertices have different colors.

When $E = \{0, 1\}$ the set A can be chosen so that to represent the configurations where no two occupied sites are adjacent (that is $i \sim j \Rightarrow x(i)x(j) \neq 1$). In statistical physics, this model is often referred as the hard-core model.

In this context, $I - i$ represents the set of all vertices that differs from the vertex $i \in I$. Given some coloring x_{I-i} of these vertices $I - i$, the set $A_i(x_{I-i})$ coincides with the set of colors $k \in E$ not appearing in the neighborhood of the vertex i.

When the reference measure λ is given by the product counting measure on the set of colors, the Markov transition $L_{I-i,i}(x_{I-i}, dx_i)$ amounts in choosing uniformly at random some color k that doesn't appear in the neighborhood of the vertex i and set $x_i = k$.

The Gibbs sampler associated with the Markov transition (8.57) is defined by a Markov chain $X_n = (X_n(i))_{i \in I} \in S = E^I$: At time n, we choose randomly a vertex $i \in I$ and we set $X_{n+1}(j) = X_n(j)$ for any $j \in I - i$. Finally $X_n(i)$ is an uniform r.v. on the set $E - X_n(N(i))$, where $N(i)$ stands for the set of all neighbors j of i (that is the vertices j s.t. $(i, j) \in \mathcal{E}$).

8.5.5 Subset sampling

We let λ be a reference probability measure on the set $S = E^I$ discussed in section 8.5.4. We assume that λ satisfies the desintegration property

$$\lambda(dx) = \lambda_{I-i}(dx_{I-i}) \ P_{I-i,i}(x_{I-i}, dx_i)$$

for some Markov transitions $P_{I-i,i}$ from E^{I-i} into E. For finite state spaces E, we can consider the product counting measures

$$\lambda(x) = \prod_{i \in I} \lambda_i(x_i) \quad \text{with} \quad \lambda_i(x_i) = \frac{1}{\mathrm{Card}(E)}$$

In this situation, we have

$$\lambda_{I-i}(dx_{I-i}) = \prod_{j \in I} \lambda_j(x_j) \quad \text{and} \quad P_{I-i,i}(x_{I-i}, dx_i) = \lambda_i(x_i)$$

We let π be the Boltzmann-Gibbs measure associated with some subset $A \subset S = E^I$ and defined by

$$\pi(dx) = \frac{1}{\lambda(A)} \, 1_A(x) \, \lambda(dx)$$

For each $i \in I$, we let A_{I-i} be the projection of the set A into the set E^{I-i} defined by the set of mappings $x_{I-i} \in E^{I-i}$ that can be extended to some mapping $x \in A$ by choosing some $k \in E$ and setting $x(i) = k$. In this slightly abusive notation, we let

$$A_i(x_{I-i}) = \{k \in E \ : \ x \in A\}$$

By construction, we have

$$1_A(x) = 1_{A_{I-i}}(x_{I-i}) \times 1_{A_{I-i}(x_{I-i})}(x_i)$$

and therefore

$$\pi(dx) \propto \underbrace{1_{A_{I-i}}(x_{I-i}) \, \lambda_{I-i}(dx_{I-i})}_{\propto \, \pi_{I-i}(dx_{I-i})} \times \underbrace{P_{I-i,i}(x_{I-i}, dx_i) \, 1_{A_{I-i}(x_{I-i})}(x_i)}_{\propto \, L_{I-i,i}(x_{I-i}, dx_i)}$$

The Gibbs sampler associated with the Markov transition (8.57) is defined by a Markov chain $X_n = (X_n(i))_{i \in I} \in S = E^I$.

At time n, we choose randomly a vertex $i \in I$ and we set $X_{n+1}(j) = X_n(j)$ for any $j \in I - i$. Finally $X_n(i)$ is a r.v. with the distribution $P_{I-i,i}(x_{I-i}, dx_i)$ restricted to the set $A_{I-i}(x_{I-i})$.

Part III

Nonlinear Monte Carlo methods

Chapter 9

Nonlinear Markov processes

9.1 Discrete time models

9.1.1 Markov-McKean models

In Section 4 and Section 5, we have presented a class of Markov chain models, with a linear evolution of the distributions of the random states. In more general situations, the law of the random states $\eta_n = \text{Law}(X_n)$ of the stochastic process satisfies a nonlinear evolution equation of the following form

$$\eta_n = \Phi_n(\eta_{n-1}) \tag{9.1}$$

for some nonnecessarily linear one step mapping Φ_n from $\mathcal{P}(E_{n-1})$ into $\mathcal{P}(E_n)$. Notice that the Markov chain model discussed in (4.14) is associated with the linear mapping

$$\Phi_n \ : \ \eta \in \mathcal{P}(E_{n-1}) \ \mapsto \ \Phi_n(\eta) = \eta K_n \in \mathcal{P}(E_n)$$

Inversely, *any* evolution equation of the form (9.1) can be interpreted as the evolution of the distributions of the random states of a Markov chain taking values in some state spaces E_n. More precisely, there always exists some (non unique) collection of Markov transition $K_{n,\eta}$ indexed by the time parameter $n \in \mathbb{N}$, and the set of probability measures η on E_{n-1}, such that

$$\eta_n = \eta_{n-1} K_{n,\eta_{n-1}} \tag{9.2}$$

For instance, we can take $K_{n,\eta}(x_{n-1}, .) = \Phi_n(\eta)$. In this situation, the random states X_n form a sequence of independent random variables with distributions η_n satisfying the evolution Equation (9.1). We refer the reader to Section 11.2.3, for applications of these models in the context of linear Gaussian filtering models, and their mean field ensemble Kalman filters.

By construction, the random states X_n of the system can again be interpreted as the distribution of "memoryless" Markov processes with elementary transitions

$$K_{n,\eta_{n-1}}(x_{n-1}, dx_n) = \mathbb{P}\left(X_n \in dx_n \mid X_{n-1} = x_{n-1}\right)$$

that depend on the distribution $\eta_{n-1} = \text{Law}(X_{n-1})$ of the current state X_{n-1} of the system. These stochastic processes are called *the McKean interpretations* of the nonlinear equation in distribution space (9.1). The distributions of the random trajectories $(X_p)_{0 \leq p \leq n}$ of the chain are given by the McKean distributions

$$\mathbb{P}\left((X_0, \ldots, X_n) \in d(x_0, \ldots, x_n)\right) = \eta_0(dx_0) \prod_{1 \leq p \leq n} K_{p,\eta_{p-1}}(x_{p-1}, dx_p) \tag{9.3}$$

The existence of an overall distribution $\mathbb{P}$ on the set of trajectories $\Omega := \prod_{n \geq 0} E_n$ of the McKean-Markov chain is granted by the Ionescu-Tulcea's theorem (cf. for instance the theorem on page 249, in the textbook by A. Shiryaev [539]).

In computational physics, and more particularly in fluid mechanics, nonlinear stochastic processes of the form (9.3) represent the evolution of complex stochastic systems with so many particles that they can be treated as a continuum. In this interpretation, a single particle interacts with the distribution of the whole population of individuals.

Another central idea, commonly used in this stochastic modeling, is to express the evolution of macroscopic quantities in terms of averages w.r.t. the distribution of microscopic variables. In kinetic theory, these stochastic models are used to represent the evolutions of various physical systems, such as gases, macroscopic fluid models, and molecular chaotic systems.

The foundations of kinetic theory were laid down by J.C. Maxwell [440, 441, 442]. Further details on these models can be found in the series of articles [119, 120, 121, 136, 308, 419, 576] and in the more probabilistic treatments [309, 445, 446, 447, 559, 566].

More recently, these Vlasov type equations have also come to play an important role in mathematical biology, as well as in socio-economics and stochastic games theory.

In biology, they are used to model animal competitions as well as bacteria dynamics with group interactions, or complex cell motions [46, 47, 459, 494]. In stochastic games theory, McKean-Vlasov models represent competing multiple class agents weakly coupled w.r.t. their dynamics and their performance [85, 157, 301, 325, 346, 443, 418, 398, 399, 400, 401].

We quote recent applications of Markov-McKean models in filtering problems arising in turbulent fluid mechanics and weather forecasting prediction developed by Ch. Baehr and his co-authors in a series of articles [27, 28, 30, 31, 407, 505, 558].

The time evolution of these stochastic kinetic models can be defined in terms of discrete generation nonlinear Markov chains discussed in Section 9.1, or in terms of the integro-differential equations presented in Section 9.2. In this connection, we also quote the series of articles [14, 64, 65, 66, 389] in the late 1990s on the rate of convergence of some time discretization schemes of McKean-Vlasov diffusions.

9.1.2 McKean-Vlasov type models

Prototypes of discrete generation and nonlinear Markov-McKean models are given by McKean-Vlasov-Fokker-Planck diffusion type models arising in fluid mechanics, as well as in mean field game theory. These models are also used in advanced signal processing, and more particularly in forecasting data assimilation problems. We refer the reader to Section 11.2.3, dedicated to a McKean diffusion type interpretation of the Kalman filter, and their mean field ensemble Kalman filters.

In dimension $d = 1$, these nonhomogeneous Markov models are given by an $\mathbb{R}$-valued stochastic process defined by the recursive equation

$$X_n - X_{n-1} = \mathbf{a_n}(X_{n-1}, \eta_{n-1}) + \boldsymbol{\sigma_n}(X_{n-1}, \eta_{n-1})\ W_n \tag{9.4}$$

with $\eta_{n-1} := \mathrm{Law}(X_{n-1})$. In the above displayed formula, X_0 is a r.v. $(W_n)_{n\geq 0}$ is a collection of i.i.d. centered Gaussian random variables with unit variance, and the drift and diffusion functions are defined by

$$\begin{aligned} \mathbf{a_n}(X_{n-1}, \eta_{n-1}) &= \int a_n(X_{n-1}, x_{n-1})\ \eta_{n-1}(dx_{n-1}) \\ \boldsymbol{\sigma_n}(X_{n-1}, \eta_{n-1}) &= \int \sigma_n(X_{n-1}, x_{n-1})\ \eta_{n-1}(dx_{n-1}) \end{aligned} \tag{9.5}$$

for some regular mappings a_n and σ_n. Whenever $\sigma_n(x, y) \geq \epsilon$, for some $\epsilon > 0$, the laws of the random states $\eta_n = \mathrm{Law}(X_n)$ satisfy the evolution Equation (9.2), with the McKean transitions given by

$$K_{n,\eta}(x, dy) = \frac{1}{\sqrt{2\pi \boldsymbol{\sigma}_{\boldsymbol{n}}^{\mathbf{2}}(x, \eta)}}\ \exp\left\{-\frac{1}{2}\left(\frac{(y - x) - \mathbf{a_n}(x, \eta)}{\boldsymbol{\sigma_n}(x, \eta)}\right)^2\right\} dy$$

9.1.3 Nonlinear state space models

A more general class of McKean-Markov chain models on some measurable state spaces E_n are given by the recursive formulae

$$X_n = F_n(X_{n-1}, \eta_{n-1}, W_n) \quad \text{with} \quad \eta_{n-1} := \text{Law}(X_{n-1}) \tag{9.6}$$

In the above display, W_n is a collection of independent, (and independent of $(X_p)_{0\leq p<n}$) r.v. taking values in some state space $\mathcal{W}_n$, and F_n is some measurable mapping from $(E_{n-1} \times \mathcal{P}(E_{n-1}) \times \mathcal{W}_n)$ into E_n.

Here again, we easily check that the laws of the random states $\eta_n = \text{Law}(X_n)$ satisfy the evolution Equation (9.2), with the McKean transitions

$$K_{n,\eta}(f)(x) = \mathbb{E}\left[f(F_n(x, \eta, W_n))\right]$$

More general discrete time generation Markov-McKean models of the form (9.6) are discussed in the series of articles [153, 158, 161, 177, 170, 208].

9.1.4 Feynman-Kac models

We consider the Feynman-Kac measures (γ_n, η_n) discussed in section 4.2.1 and for any $f_n \in \mathcal{B}_b(E_n)$ defined by

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f_n) := \mathbb{E}\left(f_n(X_n) \prod_{0\leq p<n} G_p(X_p)\right) \tag{9.7}$$

We recall that

$$\begin{array}{rcl} \gamma_{n+1}(f_{n+1}) = \gamma_n(G_n M_{n+1}(f_{n+1})) \Rightarrow \eta_{n+1}(f_{n+1}) & = & \frac{\gamma_n(G_n M_{n+1}(f_{n+1}))}{\gamma_n(G_n)} = \frac{\gamma_n(G_n M_{n+1}(f_{n+1}))/\gamma_n(1)}{\gamma_n(G_n)/\gamma_n(1)} \\ & = & \frac{\eta_n(G_n M_{n+1}(f_{n+1}))}{\eta_n(G_n)} = \Psi_{G_n}(\eta_n)\left(M_{n+1}(f_{n+1})\right) \end{array}$$

This shows that the flow of measures η_n satisfies a nonlinear evolution equation of the form (9.1), with the one step mappings

$$\Phi_{n+1}(\eta) = \Psi_{G_n}(\eta) M_{n+1} \tag{9.8}$$

In the above displayed formula, $\Psi_{G_n}(\eta)$ stands for the Boltzmann-Gibbs measures defined in (7). We recall from (8) that $\Psi_{G_n}(\eta)$ can be interpreted as a nonlinear Markov transport model

$$\Psi_{G_n}(\eta) = \eta S_{\eta,G_n} \tag{9.9}$$

for some Markov transitions S_{μ,G_n} from E_n into itself. In this situation, the one step mapping Φ_{n+1} can be rewritten as follows

$$\Phi_{n+1}(\eta) = \Psi_{G_n}(\eta) M_{n+1} = \eta K_{n+1,\eta} \quad \text{with} \quad K_{n+1,\eta} := S_{\eta,G_n} M_{n+1}$$

For instance, for $[0,1]$-valued potential functions G_n, and any $\epsilon \in [0,1]$, we can choose

$$S_{\eta,G_n}(x, dy) := \epsilon\ G(x)\ \delta_x(dy) + (1 - \epsilon\ G(x))\ \Psi_{G_n}(\eta)(dy)$$

In this situation, $K_{n+1,\eta}$ takes the form

$$K_{n+1,\eta}(x, dy) = \epsilon G_n(x)\ M_{n+1}(x, dy) + (1 - \epsilon G_n(x))\ \Psi_{G_n}(\eta) M_{n+1}(dy) \tag{9.10}$$

In continuation to the remarks we made in the beginning of Section 5, discrete time Feynman-Kac models also encapsulate without further work continuous time models. More precisely, let us consider a continuous time Markov process $(X'_t)_{t\geq 0}$ taking values in some Polish state space E, and a sequence of measurable and bounded functions

$$V \ : \ (t,x) \in (\mathbb{R}_+ \times E) \mapsto V_t(x) \in \mathbb{R}$$

As in Section 5, we consider a time mesh $(t_n)_{n\geq 0}$, with $t_0 = 0$, and we let $E_n = D([t_n, t_{n+1}], E)$ be the set of càdlàg paths from the interval $[t_n, t_{n+1}]$ into E. We consider the Feynman-Kac model (4.17) associated with the sequence of random variables X_n, and the potential functions G_n on E_n, defined by

$$X_n = (X'_t)_{t_n\leq t\leq t_{n+1}} \quad \text{and} \quad G_n(X_n) = \exp\left(\int_{t_n}^{t_{n+1}} V_t(X'_t)\ dt\right) \tag{9.11}$$

By construction, it is readily checked that

$$\mathbb{Q}_n(dx) = \frac{1}{\mathcal{Z}_n}\ \exp\left\{\int_0^{t_n} V_t(x_s)\ ds\right\}\ \mathbb{P}'_{t_{n+1}}(dx)$$

where $\mathbb{P}'_{t_{n+1}}$ is the distribution of $(X'_t)_{0\leq t\leq t_{n+1}}$ on $D([0, t_{n+1}], E)$. In the above display, $dx = d(x_s)_{s\leq t}$ stands for an infinitesimal neighborhood of the path $x = (x_s)_{s\leq t_n} \in D([0, t_{n+1}], E)$.

9.1.5 Markov chain Monte Carlo with recycling

We consider a collection of Boltzmann-Gibbs measures μ_n are defined in terms of some reference measure λ on some abstract measurable state space, say E, weighted by some product of potential functions $h_p \ : \ E \mapsto [0,\infty[$, with $p \leq n$. More formally, these measures are defined, up to some normalizing constant $\mathcal{Z}_n$, by the following formulae

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n}\ \left\{\prod_{0\leq p\leq n} h_p(x)\right\}\ \lambda(dx) \tag{9.12}$$

for some normalizing constant $\mathcal{Z}_n > 0$.

For instance, let us consider a nondecreasing sequence of inverse temperature parameters β_n, with $\beta_0 = 0 = \beta_{-1}$, and a given energy type function V, on some state space E, equipped with some reference distribution λ.

$$h_n(x) = \exp\left(-\left(\beta_n - \beta_{n-1}\right) V(x)\right)$$

in (9.12), then we find that

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n}\ \exp\left(-\beta_n V(x)\right)\ \lambda(dx) \quad \text{with} \quad \mathcal{Z}_n = \int \lambda(dx) \exp\left(-\beta_n V(x)\right)$$

If we choose the indicator functions $h_n := 1_{A_n}$, with nonincreasing sequence of subsets $A_n \subset E$, take the following simple form

$$\mu_n(dx) = \frac{1}{\mathcal{Z}_n}\ 1_{A_n}(x)\ \lambda(dx) \quad \text{and} \quad \mathcal{Z}_n := \lambda(A_n) \tag{9.13}$$

These measure restriction models arise in a variety of application domains, including in rare event simulation and stochastic optimization models. In this section, we illustrate these models in the context of calibration problems and uncertainty propagations in complex numerical codes (or in numerical meta-models). These problems are often formulated in terms of a classical input-output transformation $I \mapsto O = C(I)$. The inputs I have some distribution λ. They represent the sources of randomness, some

tuning parameters, or unknown kinetic parameters of the code. The output variable O can be interpreted as the outputs of a numerical approximation of some partial differential equation representing a given physical, chemical, or some biological phenomenon.

The prototype of questions arising in practice is the following: We are given a decreasing sequence of unlikely critical domains, say $\mathcal{O}_n$, in the space of the outputs, and we want to estimate both the probability that the outputs fall into these sets, as well as the distribution of the inputs leading to these critical events. More formally, if we set $A_n := C^{-1}(\mathcal{O}_n)$ these probabilistic objects coincide with the normalizing constants $\mathcal{Z}_n$, and the conditional distributions μ_n introduced in (9.13).

We further assume that we have a dedicated MCMC elementary transition M_n, with the target measure $\mu_n = \mu_n M_n$, at any time $n \geq 0$.

For instance, In the example (9.13) discussed above, we can choose in (9.14) the MCMC transition

$$M_n(x, dy) = K(x, dy)\ 1_{A_n}(y) + (1 - K(x, A_n))\ \delta_x(dy)$$

for any λ-reversible Markov transition K. In engineering and scientific computing literature, the mean field models associated with these restriction models are also called subset simulation algorithms [22, 23, 24, 413, 416].

By construction, it is now readily checked that

$$\left.\begin{array}{rcl} \mu_n & = & \mu_n M_n \\ \mu_n & = & \Psi_{h_n}(\mu_{n-1}) \end{array}\right\} \Longrightarrow \mu_n = \mu_n M_n = \Psi_{h_n}(\mu_{n-1}) M_n \tag{9.14}$$

In other words, if we set $G_n = h_{n+1}$, in the Feynman-Kac model (4.17), then we have $\mu_n = \eta_n$. In other words, the Boltzmann-Gibbs measures μ_n can also be interpreted as the n-th time marginals of the Feynman-Kac measures $\mathbb{Q}_n$, with the potential functions $G_n = h_{n+1}$ and the reference Markov chain X_n, associated with the Markov chain Monte Carlo elementary transition M_n. It is also easily checked that

$$\mathcal{Z}_n = \mu_{n-1}(h_n) \times \mathcal{Z}_{n-1} \Rightarrow \mathcal{Z}_n / \mathcal{Z}_0 = \mathbb{E}\left(\prod_{0 \leq p \leq n} G_p(X_p)\right) \tag{9.15}$$

In the above display, X_n stands for a reference Markov chain with initial distribution μ_0 and elementary transition probabilities M_n.

9.2 Continuous time models

To underline the role of mean field simulation in the numerical solving of nonlinear integro-differential equations, we provide a description of some continuous time versions of the Markov transport Equation (9.2). The discrete time approximation of these models are presented in section 9.2.5.

9.2.1 Nonlinear jump-diffusion models

We consider the general jump-diffusion processes (5.3) discussed in the end of Section 4. Continuous time McKean processes X'_t are defined as in (5.1) and (5.3) by replacing the parameters $(a'_t, \sigma'_t, \lambda'_t, P'_t)$ by some collection of functional parameters

$$(a'_t(.,\eta'_t), \sigma'_t(.,\eta'_t), \lambda'_{t,.,\eta'_t}, P'_{t,\eta'_t})$$

that depend on the distribution flow $\eta'_t = \mathrm{Law}(X'_t)$ of the random states X'_t. To avoid unnecessary technical discussion, it is implicitly assumed that the model parameters are sufficiently regular, so that the corresponding nonlinear jump diffusion process is well defined on the real line $\mathbb{R}_+$.

In this situation, the corresponding integro-differential Equation (5.43) takes the following form

$$\frac{d}{dt}\eta'_t(f) = \eta'_t\left(L_{t,\eta'_t}(f)\right) \tag{9.16}$$

with the collection of integro-differential operators, indexed by the time parameter t, and the set of probability distributions η on $\mathbb{R}^d$, given by

$$\begin{aligned} L_{t,\eta}(f)(x) &= L'_{t,\eta}(f)(x) + \lambda'_{t;\eta}(x) \int [f(y) - f(x)]\ P'_{t,\eta}(x, dy) \\ L'_{t,\eta} &:= \sum_{i=1}^{d} a'_{t,i}(x,\eta)\ \partial_i + \frac{1}{2}\sum_{i,j=1}^{d} \left(\sigma'_t(x,\eta)(\sigma'_t(x,\eta))^T\right)_{i,j}\ \partial_{i,j} \end{aligned} \tag{9.17}$$

We notice that the second order differential operator $L'_{t,\eta}$ defined in (9.17) is the infinitesimal generator of the nonlinear diffusion process

$$dX'_t = a'_t(X'_t, \eta'_t)\ dt + \sigma'_t(X'_t, \eta'_t)\ dW_t \tag{9.18}$$

The corresponding nonlinear diffusion jump process X'_t evolves between jumps times T_n as in (9.18). As in (5.3), the jump times T_n are defined by the recursion

$$T_n = \inf\left\{ t \geq T_{n-1}\ :\ \int_{T_{n-1}}^{t} \lambda'_{u,\eta'_u}(X'_u)\ du \geq e_n \right\} \tag{9.19}$$

At jump times T_n, the process at X'_{T_n-} jumps to a new location X'_{T_n} randomly chosen with a distribution $P'_{T_n-,\eta'_{T_n-}}(X'_{T_n-}, dy)$ that depends on the distribution η'_{T_n-} of the random state X'_{T_n-} of the process before the jump.

9.2.2 Nonlinear integro-differential equations

We further assume that the jump transition $P'_{t,\eta'_t}(x, dy) = q_{t,\eta'_t}(x, y)\ dy$ and the law of the random states $\eta'_t(dy) = p_t(y)\ dy$ of the jump-diffusion process (9.18) have a smooth density $q_{t,\eta'_t}(x, y)$, and $p_t(y)$ w.r.t. the Lebesgue measure dy on $\mathbb{R}^d$. To simplify the presentation, with a slight abuse of notation we set $\left(a'_t(.,p_t), \sigma'_t(.,p_t), q_{t,p_t}, \lambda'_{t,p_t}, L'_{t,p_t}\right)$ instead of $\left(a'_t(.,\eta'_t), \sigma'_t(.,\eta'_t), q_{t,\eta'_t}, \lambda'_{t,\eta'_t}, L'_{t,\eta'_t}\right)$.

In this notation, the equation (9.16) takes the form

$$\begin{aligned} \frac{\partial}{\partial t}\eta'_t(f) &= \int f(x)\ \frac{\partial p_t}{\partial t}(x)\ dx \\ &= \int p_t(x)\ L'_{t,p_t}(f)(x)\ dx \\ &\qquad + \int p_t(x)\ \lambda'_{t,p_t}(x)\ q_{t,p_t}(x,y)\ (f(y) - f(x))\ dxdy \end{aligned}$$

from which we conclude that

$$\frac{\partial p_t}{\partial t}(x) = L'^{\star}_{t,p_t}(p_t)(x) + \left(\int p_t(y)\ \lambda'_{t,p_t}(y)\ q_{t,p_t}(y,x)\ dy\right) - p_t(x)\ \lambda'_t(x) \tag{9.20}$$

with the dual differential operator

$$\begin{aligned} &L'^{\star}_{t,p_t}(p_t) \\ &= -\sum_{i=1}^{d}\ \partial_i\left(a'_{t,i}(.,p_t)\ p_t\right) + \tfrac{1}{2}\ \sum_{i,j=1}^{d}\ \partial_{i,j}\left(\left(\sigma'_t(.,p_t)(\sigma'_t(.,p_t))^T\right)_{i,j}\ p_t\right) \end{aligned}$$

Inversely, any equation of the form (9.20) can be interpreted as the probability densities of the random states of a jump diffusion model of the form (9.19).

The mean field particle methods developped in Section 9.3.2, and in Section 9.3 allow to reduce the numerical solving of nonlinear integro-differential equations of the form (9.20) to the stochastic simulation of the jump-diffusion process (9.18). The nonlinear equations (9.16) and (9.20), and their probabilistic interpretations where introduced in the early 1950s in [368] by M. Kac to analyze the Vlasov kinetic equation of plasma [395]. These lines of ideas were further developed by H. P. McKean in [445].

For a more thorough and rigorous discussion on these nonlinear models, and their mean field particle interpretations we refer the reader to the series of articles by C. Graham [316, 317, 318], and the ones by C. Graham and S. Méléard [309, 311, 313, 314, 315].

9.2.3 Normalized Feynman-Kac flows

When the functions $a'_{t,i}(x,\eta) = a'_{t,i}(x)$ and $\sigma'_t(x,\eta) = \sigma'_t(x)$ doesn't depend on the parameter η, the operator $L'_{t,\eta} = L'_t$ defined in (9.17) coincides with the infinitesimal generator of the diffusion process defined in (5.1). We also consider the parameters

$$\lambda'_{t;\eta} := V_t(x) \quad \text{and} \quad P'_{t,\eta}(x,dy) := \eta(dy) \tag{9.21}$$

In this situation, the Equation (9.16) is given by the following quadratic evolution model

$$\partial_t \eta'_t(f) = \eta'_t\left(L'_t(f)\right) + \eta'_t(V_t)\eta'_t(f) - \eta'_t(fV_t) = \eta'_t\left(L^V_{t,\eta'_t}(f)\right) \tag{9.22}$$

with the "normalized" Schrödinger operator

$$L^V_{t,\eta} := L'_t - (V_t - \eta_t(V))$$

Using (5.40) the solution of this equation is given by

$$\eta'_t(f) = \gamma'_t(f)/\gamma'_t(1)$$

where γ'_t is the unnormalized Feynman-Kac measure defined by

$$\begin{aligned}
\gamma'_t(f) &= \mathbb{E}\left[f(X'_t)\ \exp\left(-\int_0^t V_u(X'_u)du\right)\right] \\
&= \int \eta'_0(dx_0)\ \mathbb{E}\left[f(X'_t)\ \exp\left(-\int_s^t V_u(X'_u)du\right)\ |\ X'_0 = x_0\right] \\
&= \int \eta'_0(dx_0)\ P^V_{0,t}(f)(x_0) = \int \gamma'_s(dx_s)\ P^V_{s,t}(f)(x_s)
\end{aligned}$$

with the Feynman-Kac semigroup defined in (5.33) given by

$$P^V_{s,t}(f)(x_s) = \mathbb{E}\left[f(X'_t)\ \exp\left(-\int_s^t V_u(X'_u)du\right)\ |\ X'_s = x_s\right]$$

In the above displayed formulae, X'_t stands for the diffusion process defined in (5.1) with infinitesimal generator L'_t.

These continuous time models are discussed in [170, 185, 187, 188, 194], as well as in [309, 445, 446, 447, 448].

Returning to the McKean interpretation models discussed in Section 9.2.1, the flow of measures $\eta'_t = \mathrm{Law}(X_t)$ can be interpreted as the distributions of the random states X_t of a jump type Markov process.

Between the jumps, X_t follows the diffusion Equation (5.1). At jump times T_n, occurring with the stochastic rate $V_t(X_t)$, the process $X_{T_n-} \rightsquigarrow X_{T_n}$ jumps to a new location, randomly chosen with the distribution $P'_{T_n-,\eta_{T_n-}}(x,dy) := \eta'_{T_n-}(dy)$.

9.2.4 A jump type Langevin model

We consider an non homogeneous overdamped Langevin diffusion, on an energy landscape associated with a given energy function $V \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}_+)$ on $E = \mathbb{R}^d$, for some $d \geq 1$. This model is defined by the following diffusion equation

$$dX'_t = -\beta_t \ \nabla V(X'_t) + \sqrt{2}\ dB_t$$

where ∇V denotes the gradient of V, β an inverse temperature parameter, and B_t a standard Brownian motion on $\mathbb{R}^d$ (also called a Wiener process). The infinitesimal generator associated with this continuous time process is given by the second order differential operator

$$L'_{\beta_t} = -\beta_t \ \nabla V \cdot \nabla + \triangle$$

Under some regularity conditions on V, for any fixed $\beta_t = \beta$ the diffusion X'_t is geometrically ergodic with an invariant measure given by

$$d\pi_\beta = \frac{1}{\mathcal{Z}_\beta}\ e^{-\beta V}\ d\lambda \tag{9.23}$$

where λ stands for the Lebesgue measure on $\mathbb{R}^d$, and $\mathcal{Z}_\beta$ is a normalizing constant. When the inverse temperature parameter β_t depends on the time parameter t, the time inhomogeneous diffusion X'_t has a time inhomogeneous generator L'_{β_t}.

We further assume that $\pi_{\beta_0} = \mathrm{Law}(X'_0)$, and we set $\beta'_t := \frac{d\beta_t}{dt}$. By construction, we have

$$\frac{d}{dt}\pi_{\beta_t}(f) = \beta'_t\ (\pi_{\beta_t}(V)\pi_{\beta_t}(f) - \pi_{\beta_t}(fV)) \quad \text{and} \quad \pi_{\beta_t} L'_{\beta_t} = 0$$

Using these observations, we readily check that π_{β_t} satisfies the Feynman-Kac evolution equation defined as in (9.22) by replacing V_t by $\beta'_t V$. More formally, we have that

$$\frac{d}{dt}\pi_{\beta_t}(f) = \pi_{\beta_t}\left(L'_{\beta_t}(f)\right) + \beta'_t\,(\pi_{\beta_t}(V)\pi_{\beta_t}(f) - \pi_{\beta_t}(fV))$$

from which we conclude that

$$\pi_{\beta_t}(f) = \gamma_t(f)/\gamma_t(1) \quad \text{with} \quad \gamma_t(f) := \mathbb{E}\left(f(X'_t)\ \exp\left(-\int_0^t \beta'_s\ V(X'_s)ds\right)\right)$$

It is also easily checked that

$$\gamma_t(1) := \mathbb{E}\left(\exp\left(-\int_0^t \beta'_s\ V(X'_s)ds\right)\right) = \exp\left(-\int_0^t \beta'_s\ \eta_s(V)ds\right) = \mathcal{Z}_{\beta_t}/\mathcal{Z}_{\beta_0}$$

This formula is known as the Jarzinsky equality [137, 138, 359, 360]. In statistical physics, the weight functions

$$\mathcal{V}_t(X') = \int_0^t \beta'_s\ V(X'_s)\ ds$$

represent the out of equilibrium virtual work of the system on the time horizon t.

In summary, we have described a McKean interpretation of Boltzmann-Gibbs measures (9.23) associated with some nondecreasing inverse cooling schedule. Returning to the McKean interpretations of the Feynman-Kac models discussed in Section 9.2.3, the flow of measures

$$\eta_t := \pi_{\beta_t} = \mathrm{Law}(X_t)$$

can be interpreted as the distributions of the random states X_t of a jump type Markov process. Between the jumps, X_t follows the Langevin diffusion Equation (7.32). At jump times T_n, with the stochastic rate $\beta'_t V_t(X_t)$, the process $X_{T_n-} \rightsquigarrow X_{T_n}$ jumps to a new site randomly chosen with the distribution $\eta_{T_n-}(dy)$.

9.2.5 Time discretization schemes

Using the same reasoning as in the end of Section 5, a discrete time approximation on a time mesh t_n of the jump-diffusion model discussed in Section 9.2.1 is given by a Markov chain X_n with elementary transitions

$$K_{n+1,\eta_n}(x,dz) = \int M_{n+1,\eta_n}(x,dy)\ J_{n+1,\overline{\eta}_n}(y,dz) \tag{9.24}$$

with the probability measures $\overline{\eta}_n$ given by

$$\overline{\eta}_n = \eta_n \qquad \text{or} \qquad \overline{\eta}_n := \eta_n M_{n+1,\eta_n}$$

In the above displayed integral formula, M_{n+1,η_n} stands for the transition of the Markov chain

$$X_{n+1} = X_n + \mathbf{a_{n+1}}\left(X_n,\eta_n\right)\ \Delta + \boldsymbol{\sigma}_{n+1}\left(X_n,\eta_n\right)\ \left(W_{t_{n+1}} - W_{t_n}\right)$$

with $(\mathbf{a}_{n+1},\boldsymbol{\sigma}_{n+1}) = \left(a'_{t_n},\sigma'_{t_n}\right)$, and $\eta_n = \text{Law}(X_n)$, The geometric jump type Markov transition $J_{n+1,\eta}$ is now defined in terms of the parameters

$$(G_{n+1,\eta},P_{n+1,\eta}) = \left(\exp\left(-\Delta\lambda_{t_{n+1},\eta}\right),P'_{t_{n+1},\eta}\right)$$

by the following equation

$$J_{n+1,\eta}(y,dz) = G_{n+1,\eta}(y)\ \delta_y(dz) + (1 - G_{n+1,\eta}(y))\ \ P_{n+1,\eta}(y,dz)$$

In this situation, the discrete generation model (9.2), associated with the Markov transitions (9.24), can be seen as an approximation of the jump type Markov process with infinitesimal generator (9.16). Therefore, using the same line of arguments as the ones given in Section 5.6, the distribution flow η_n of the random states of the discrete generation process given in (9.2) approximates the solution η'_{t_n} of the continuous time integro-differential Equation (9.16). More precisely, taking $\Delta \simeq 0$ in (9.24), we have the approximation

$$\begin{aligned} J_{n+1,\overline{\eta}_n} - Id \quad &= \left(1 - \exp\left(-\Delta\lambda_{t_{n+1},\overline{\eta}_n}\right)\right)\ \left(P'_{t_{n+1},\overline{\eta}_n} - Id\right) \\ &\simeq \quad \lambda_{t_{n+1},\overline{\eta}_n}\ \left(P'_{t_{n+1},\overline{\eta}_n} - Id\right) \simeq \lambda_{t_{n+1},\eta_n}\ \left(P'_{t_{n+1},\eta_n} - Id\right)\ \Delta \end{aligned}$$

and for any smooth functions f we also have that

$$\begin{aligned} &(M_{n+1,\eta_n} - Id)(f)(x) \\ &\simeq \mathbb{E}\left(f\left(x + a'_{t_n}\left(x,\eta_n\right)\ \Delta + \sigma'_{t_n}\left(x,\eta_n\right)\ \left(W_{t_{n+1}} - W_{t_n}\right)\right) - f(x)\right) \\ &\simeq L'_{t_n,\eta_n}(f)(x)\ \Delta \end{aligned}$$

with the second order differential operator $L'_{t,\eta}$ defined in (9.17). As in (5.14), we conclude that

$$K_{n+1,\eta_n} \simeq Id + L_{t_n,\eta_n}\ \Delta \simeq Id + L_{t_n,\eta'_{t_n}} \tag{9.25}$$

with the infinitesimal generator $L_{t,\eta}$ defined in (9.17).

If we choose $\overline{\eta}_n = \eta_n M_{n+1,\eta_n}$ in (9.24), then the elementary transitions of the McKean-Markov chain $X_n \rightsquigarrow X_{n+1}$ are decomposed into two steps

$$X_n \xrightarrow{\quad M_{n+1,\eta_n}\quad} \overline{X}_n \xrightarrow{\quad J_{n+1,\overline{\eta}_n}\quad} X_{n+1} \quad \text{with} \quad \text{Law}\left(\overline{X}_n\right) = \overline{\eta}_n$$

If we set $X_{n+1/2} := \overline{X}_n$, and $\eta_{n+1/2} := \overline{\eta}_n$, then the corresponding discrete generation McKean model is defined in terms of the slightly different equations

$$\eta_n \rightsquigarrow \eta_{n+1/2} = \eta_n K_{n+1/2,\eta_n} \rightsquigarrow \eta_{n+1} = \eta_{n+1/2} K_{n+1,\eta_{n+1/2}} \tag{9.26}$$

with the McKean transitions

$$\left(K_{n+1/2,\eta}, K_{n+1,\eta}\right) = (M_{n+1,\eta}, J_{n+1,\eta})$$

Up to a change of time, the discrete evolution model described above has exactly the same form as the one discussed in (9.2). To the best of our knowledge, the refined analysis of the couple of discrete time approximation schemes discussed above has not been covered in the literature of McKean diffusions with jumps.

9.3 Mean field particle methods

9.3.1 Discrete time models

With the exception of some very special cases, the evolution Equations (9.1) and their continuous time versions (9.16) cannot be solved explicitly. For instance, in the context of the Feynman-Kac models presented in (9.8), it can be shown that the measures η_n, or their continuous time versions (5.34) discussed in section 5.5, cannot be represented in a closed form on some finite dimensional state-space. A proof of this result can be found in the article by M. Chaleyat-Maurel and D. Michel [135], in the context of nonlinear filtering problems. In addition, conventional harmonic type approximation schemes, or related linearization style techniques, such as the extended Kalman filter, often provide poor estimation results for highly nonlinear models.

On the other hand, more traditional numerical techniques for solving evolution equations of the form (9.2) using deterministic type grid approximations require extensive calculations, and they rarely cope with nonlinear distribution flows with strongly varying probability masses, or with high dimensional problems. In contrast with these conventional numerical techniques, the mean field IPS technology presented in this section provides an accurate stochastic grid approximation scheme, equipped with an interacting mechanism tracking the probability mass variations of the distribution flow. In other words, these advanced Monte Carlo methods take advantage of the nonlinearities of the model, to drive the particle populations in regions with high probability mass.

To design a mean field IPS model associated with a given collection of McKean transitions $K_{n,\eta}$, we further assume that we have a dedicated Monte Carlo simulation tool to draw independent random samples of the elementary transitions

$$K_{n,m(y)}(x_{n-1}, dx_n) \tag{9.27}$$

for any empirical measure associated with N given states $(y^i)_{1\leq i\leq N} \in E_{n-1}^N$ given by the formula

$$m(y) := \frac{1}{N} \sum_{1\leq i\leq N} \delta_{y^i}$$

When the transitions (9.27) are easy to simulate, we design an N-particle approximation model by evolving a Markov chain $\xi_n := \left(\xi_n^i\right)_{1\leq i\leq N} \in E_n^N$ with elementary transitions

$$\mathbb{P}\left(\xi_{n+1}^{(N)} \in dx_{n+1} \mid \xi_n\right) := \prod_{1\leq i\leq N} K_{n+1,\eta_n^N}(\xi_n^{(N,i)}, dx_{n+1}^i) \tag{9.28}$$

where $dx_{n+1} = dx_{n+1}^1 \times \ldots \times dx_{n+1}^N$ stands for an infinitesimal neighborhood of a point $x_{n+1} = (x_{n+1}^i)_{1\leq i\leq N} \in E_{n+1}^N$, and η_n^N stands fo the empirical measures defined by

$$\eta_n^N := m\left(\xi_n^{(N)}\right) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^{(N,i)}}$$

The initial system $\xi_0 = (\xi_0^{(N,i)})_{1\leq i\leq N}$ consists of N i.i.d. random variables with common law η_0.

To simplify the presentation, with a slight abuse of notation, when there is no possible confusion we often suppress the index parameter $(.)^{(N)}$ *so that we write* ξ_n *and* ξ_n^i *instead of* $\xi_n^{(N,i)}$ *and* $\xi_n^{(N,i)}$.

The N-particle model ξ_n approximates the target probability distributions η_n by the occupation measures η_n^N of a large cloud of interacting particles; that is, we have that

$$\eta_n^N \longrightarrow_{N\to\infty} \eta_n$$

More is true, we can also prove that the occupation measure of the random trajectories of the particles converges, as $N \to \infty$, to the McKean measures defined in (9.3); that is, we have that

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{(\xi_0^i,\ldots,\xi_n^i)}(d(x_0,\ldots,x_n)) \longrightarrow_{N\to\infty} \eta_0(dx_0) \prod_{1\leq p\leq n} K_{p,\eta_{p-1}}(x_{p-1},dx_p) \tag{9.29}$$

In the context of Feynman-Kac models, the analysis of the McKean particle measures described above is discussed in some detail in the book [172]. To our knowledge, their analysis for general discrete generation mean field models is still an open research subject.

The local sampling errors induced by the mean field particle model are expressed in terms of the empirical random field sequence V_n^N defined by

$$V_{n+1}^N = \sqrt{N}\ \left[\eta_{n+1}^N - \Phi_{n+1}\left(\eta_n^N\right)\right]$$

Notice that V_{n+1}^N is alternatively defined by the following stochastic perturbation formulae

$$\eta_{n+1}^N = \Phi_{n+1}\left(\eta_n^N\right) + \frac{1}{\sqrt{N}}\ V_{n+1}^N \tag{9.30}$$

For $n = 0$, we also set

$$V_0^N = \sqrt{N}\ \left[\eta_0^N - \eta_0\right] \Leftrightarrow \eta_0^N = \eta_0 + \frac{1}{\sqrt{N}}\ V_0^N$$

In this interpretation, the N-particle model can also be interpreted as a stochastic perturbation of the limiting system

$$\eta_{n+1} = \Phi_{n+1}\left(\eta_n\right)$$

It is rather elementary to check that

$$\begin{aligned}
\mathbb{E}\left(V_{n+1}^N(f) \mid \xi_n\right) &= 0 \\
\mathbb{E}\left(V_{n+1}^N(f)^2 \mid \xi_n\right) &= \int \eta_n^N(dx)\ K_{n+1,\eta_n^N}\left(\left[f - K_{n+1,\eta_n^N}(f)(x)\right]^2\right)(x) \leq \mathrm{osc}(f)^2
\end{aligned}$$

9.3.2 Continuous time models

The continuous time mean field particle model associated with the integro-differential equations discussed in Section 9.2.1, Section 9.2.2, and Section 9.2.3, is a Markov chain $\xi_t^{(N)} := (\xi_t^{(N,i)})_{1\leq i\leq N}$ on the product state space $(\mathbb{R}^d)^N$, with infinitesimal generator defined, for sufficiently regular functions F on $(\mathbb{R}^d)^N$, by the following formulae

$$\mathcal{L}_t(F)(x^1,\ldots,x^N) := \sum_{1\leq i\leq N} L_{t,m(x)}^{(i)}(F)(x^1,\ldots,x^i,\ldots,x^N) \tag{9.31}$$

In the above display, $m(x) := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{x^i}$ stands for the occupation measure of the population $x = (x^i)_{1\leq i\leq N} \in E^N$; and for any $\eta \in \mathcal{P}(\mathbb{R}^d)$ $L_{t,\eta}^{(i)}$ stands for the operator $L_{t,\eta}$ defined in (9.16),

acting on the function $x^i \mapsto F(x^1, \ldots, x^i, \ldots, x^N)$. In other words, every individual $\xi_t^{(N,i)}$ is a jump-diffusion model defined as in (9.18) and (9.19) by replacing the unknown measures η'_t by their particle approximations $\eta_t'^N = \frac{1}{N}\sum_{j=1}^N \delta_{\xi_t^{(N,j)}}$.

For instance, using (9.21), the generator of the Feynman-Kac model (9.17) is given for any sufficiently regular function F by the formula

$$
\begin{aligned}
&L^{(i)}_{t,m(x)}(F)(x^1, \ldots, x^N) \\
&:= \sum_{1\leq i\leq N} L_t'^{(i)}(F)(x^1, \ldots, x^i, \ldots, x^N) + \sum_{1\leq i\leq N} V(x^i) \\
&\qquad\qquad \times \int \left[F(x^1, \ldots, x^{i-1}, y, x^{i+1}, \ldots, x^N) - F(x^1, \ldots, x^i, \ldots, x^N)\right]\ m(x)(dy)
\end{aligned}
$$

Between jumps, the particles evolve independently as the diffusion process defined in (5.1). At rate V, the particles jump to a new location randomly selected in the current population.

Using Ito's formula we have that

$$dF(\xi_t) = \mathcal{L}_t(F)(\xi_t)\ dt + d\mathcal{M}_t(F)$$

for some martingale $\mathcal{M}_t(F)$ with predictable increasing process defined by

$$\langle \mathcal{M}(F)\rangle_t := \int_0^t \Gamma_{\mathcal{L}_s}(F, F)(\xi_s)\ ds$$

We recall that the "carré du champ" operator $\Gamma_{\mathcal{L}_s}$ associated to $\mathcal{L}_s$ is defined by

$$\Gamma_{\mathcal{L}_s}(F, F)(x) := \mathcal{L}_s\left[(F - F(x))^2\right](x) = \mathcal{L}_s(F^2)(x) - 2F(x)\mathcal{L}_s(F)(x)$$

For empirical test functions of the following form

$$F(x) = m(x)(f) = \frac{1}{N}\sum_{i=1}^N f(x^i)$$

with $f \in D(L)$, we find that

$$
\begin{aligned}
\mathcal{L}_s(F)(x) &= m(x)(L_{s,m(x)}(f)) \\
\Gamma_{\mathcal{L}_s}(F, F)(x) &= \frac{1}{N}\, m(x)\left(\Gamma_{L_{s,m(x)}}(f, f)\right)
\end{aligned}
\tag{9.32}
$$

From this discussion, it should be clear that

$$\eta_t^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\xi_t^i} \implies d\eta_t^N(f) = \eta_t^N(L_{t,\eta_t^N}(f))\ dt + \frac{1}{\sqrt{N}}\, dM_t^N(f) \tag{9.33}$$

with the martingale

$$M_t^N(f) = \sqrt{N}\ \mathcal{M}_t(F) \tag{9.34}$$

The predictable angle bracket is given by

$$\langle M^N(f)\rangle_t := \int_0^t \eta_s^N\left(\Gamma_{L_{s,\eta_s^N}}(f, f)\right)\ ds$$

From the r.h.s. perturbation formulae (9.33), we conclude that η_t^N "almost solve," as $N \uparrow \infty$, the nonlinear evolution Equation (9.16). For a more thorough discussion on these continuous time models, we refer the reader to the review article [182], and the references therein.

As for the linear type jump-diffusion models, continuous time mean field models are far from being related to a concrete Monte Carlo simulation technique, since most of the time the resulting interacting jump diffusion processes cannot be sampled perfectly. One natural way to solve this problem is to consider the time discretization schemes presented in Section 9.2.5, and the discrete generation mean field models discussed in Section 9.3.

We end this section with a discussion on the connections between these discrete generation mean field models and their continuous time version discussed in Section 9.3.2. For time homogeneous state spaces $E_n = E = \mathbb{R}^d$, the Markov transition $\mathcal{M}_{n+1}$ of the discrete generation mean field model (9.28) on E^N is defined for any sufficiently regular function F on E^N by

$$\mathcal{M}_{n+1}(F)(x) := \int_{E^N} \left(\prod_{1 \leq i \leq N} K_{n+1,m(x)}(x, dy^i) \right) F(y^1, \ldots, y^N)$$

When the collection of transitions $K_{n+1,\eta}$ are given by (9.24), using the approximation formula (9.25) with $\Delta \simeq 0$ we find that

$$\begin{aligned} [\mathcal{M}_{n+1} - Id]\,(F)(x) &\simeq \prod_{1 \leq i \leq N} \left(Id + \Delta\, L_{t_n,m(x)}\right)^{(i)} F(x^1, \ldots, x^i, \ldots, x^N) - F(x) \\ &\simeq \sum_{1 \leq i \leq N} L^{(i)}_{t_n,m(x)}(F)(x^1, \ldots, x^i, \ldots, x^N)\ \Delta = \mathcal{L}_{t_n}(F)(x)\ \Delta \end{aligned}$$

In the above display, $\mathcal{L}_t$ and $L_{t,\eta}$ stands for the generators defined in (9.31) and (9.16); and $\left(Id + \Delta\, L_{t_n,m(x)}\right)^{(i)}$ stands for the operator $\left(Id + L_{t_n,m(x)}\,\Delta\right)$ acting on the function $x^i \mapsto F(x^1, \ldots, x^i, \ldots, x^N)$. This shows that the discrete generation mean field model can be interpreted as an Euler type approximation on a time mesh t_n of the continuous time particle model introduced in Section 9.3.2.

9.4 Some illustrations

9.4.1 McKean-Vlasov diffusion models

The effective sampling condition stated in (9.27) is clearly satisfied for the McKean-Vlasov models (1.24) introduced in Section 1.4.1.1. In this case (9.27) is the distribution of the random variable

$$x_{n-1} + \frac{1}{N} \sum_{1 \leq i \leq N} a_n(x_{n-1}, y^i) + \frac{1}{N} \sum_{1 \leq i \leq N} \sigma_n(x_{n-1}, y^i)\ W_n \tag{9.35}$$

In this situation, using the formula (9.35) we find that the N-particle model defined in (9.28) can be rewritten in the following form

$$\begin{aligned} \xi^i_{n+1} &:= \xi^i_n + \mathbf{a_{n+1}}(\xi^i_n, \eta^N_n) + \boldsymbol{\sigma_{n+1}}(\xi^i_n, \eta^N_n)\ W^i_{n+1} \\ &= \xi^i_n + \frac{1}{N} \sum_{1 \leq j \leq N} a_{n+1}(\xi^i_n, \xi^j_n) + \frac{1}{N} \sum_{1 \leq j \leq N} \sigma_{n+1}(\xi^i_n, \xi^j_n)\ W^i_{n+1} \end{aligned}$$

In the above displayed formulae, $\eta^N_n = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_n}$ is the occupation measure of the population at time n; and $(W^i_{n+1})_{1 \leq i \leq N}$ stands for N independent copies of the random variables W_{n+1}. Illustrations of these Gaussian mean field models in the context of filtering problems are discussed in some detail in Section 11.2.3, dedicated to ensemble Kalman filters.

9.4.2 General state space models

The effective sampling condition (9.27) is also met for the general state space models presented in (9.6), as soon as we have a dedicated Monte Carlo technique to sample the random variables $F_n(x_{n-1}, m(y), W_n)$. In this case (9.27) is given by

$$K_{n,m(y)}(x_{n-1}, dx_n) = \mathbb{P}\left(F_n(x_{n-1}, m(y), W_n) \in dx_n\right)$$

In this situation, we find that the N-particle model defined in (9.28) can be rewritten in the following form

$$\xi_{n+1}^i = F_{n+1}(\xi_n^i, \eta_n^N, W_{n+1}^i) \quad \text{with} \quad \eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \tag{9.36}$$

where $(W_{n+1}^i)_{1\leq i\leq N}$ stands for N independent copies of the r.v. W_{n+1}.

9.4.3 Interacting jump-diffusion models

We return to the Euler type approximation models discussed in Section 9.2.5. In this context, if we choose $\overline{\eta}_n = \eta_n$ in (9.24), then the mean field IPS model $\xi_n = (\xi_n^i)_{1\leq i\leq N}$ is defined by sampling N random transitions

$$\xi_n^i \rightsquigarrow \xi_{n+1}^i \sim \left(M_{n+1,\eta_n^N} J_{n+1,\eta_n^N}\right)(\xi_n^i, .) \quad \text{with} \quad \eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}$$

If we choose $\overline{\eta}_n := \eta_n M_{n+1,\eta_n}$ in (9.24), then the mean field IPS model $\xi_n = (\xi_n^i)_{1\leq i\leq N}$ associated with the distribution flow (9.26) is now defined in terms of the two step transitions

$$\xi_n^i \rightsquigarrow \xi_{n+1/2}^i \sim M_{n+1,\eta_n^N}(\xi_n^i, .) \rightsquigarrow \xi_{n+1}^i \sim J_{n+1,\eta_{n+1/2}^N}(\xi_{n+1/2}^i, .)$$

with the occupation measures at the intermediate time steps defined by

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \quad \text{and} \quad \eta_{n+1/2}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n+1/2}^i}$$

9.4.4 Feynman-Kac particle models

Condition (9.27) is also met for the Feynman-Kac model discussed in (9.10), as soon as the potential functions $G_n(x)$ can be evaluated at any state x, and as soon as we can draw independent random samples of the elementary transitions M_n. In this situation, the McKean transitions (9.27) are given by the following formulae

$$\begin{aligned} &K_{n,m(y)}(x_{n-1}, dx_n) \\ &= \epsilon G_{n-1}(x_{n-1})\; M_n(x_{n-1}, dx_n) + (1 - \epsilon G_{n-1}(x_{n-1}))\; \Psi_{G_{n-1}}\left(m(y)\right) M_n(dx_n) \end{aligned} \tag{9.37}$$

with the weighted empirical measure

$$\Psi_{G_{n-1}}\left(m(y)\right) M_n := \sum_{1\leq i\leq N} \frac{G_{n-1}(y^i)}{\sum_{1\leq j\leq N} G_{n-1}(y^j)}\; M_n(y^i, .)$$

In the above display, $\Psi_{G_{n-1}}$ stands for the Boltzmann-Gibbs transformations defined in (7).

In this situation, we use formula (9.37) to check that the N-particle model defined in (9.28) can be interpreted in terms of an interacting jump type genetic type particle model with a two step selection-mutation transition.

During the selection stage, each particle ξ^i_{n-1} evaluates its potential value $G_{n-1}(\xi^j_{n-1})$. With a probability $\epsilon G_{n-1}(\xi^i_{n-1})$ it remains in the same location. Otherwise, it jumps to a fresh new location ξ^j_{n-1} randomly chosen with a probability proportional to $G_{n-1}(\xi^j_{n-1})$. During the mutation stage, each particle evolves randomly according to the Markov transition M_n.

From the statistical, or from the stochastic, point of view, these interacting particle systems can be interpreted as a sophisticated acceptance-rejection sampling technique, equipped with an interacting recycling mechanism.

This mean field stochastic algorithm can also be interpreted as a population of individuals mimicking natural evolution mechanisms. During a mutation stage, the particles evolve independently of one another, according to the same probability transitions M_n. During the selection stage, particles with small relative values are killed, while the ones with high relative values are multiplied.

9.5 Flows of positive measures

We consider a sequence of measurable state spaces E_n and some (nonnecessarily linear) one step mapping Ξ_n from $\mathcal{M}_+(E_{n-1})$ into $\mathcal{M}_+(E_n)$. We associate with these objects the evolution equations

$$\gamma_n = \Xi_n(\gamma_{n-1})$$

starting from some initial nonnegative measure $\gamma_0 \in \mathcal{M}_+(E_0)$. These abstract models are natural extensions of the $\mathcal{P}(E_n)$-valued equations discussed in Section 9.1, to evolutions equations in the space of nonnegative measures. We observe that these evolutions equations can always be rewritten in the following form

$$\gamma_n = \gamma_{n-1} Q_{n,\gamma_{n-1}} \tag{9.38}$$

for some collection of bounded integral operators $Q_{n,\gamma}$ from $\mathcal{B}_b(E_n)$ into $\mathcal{B}_b(E_{n-1})$, indexed by the time parameter $n \geq 1$ and the set of measures $\gamma \in \mathcal{M}_+(E_n)$. As in (9.2), the choice of these operators is not unique; for instance, we can take $Q_{n,\gamma_{n-1}}(x,.) \propto \Xi_n(\gamma_{n-1})$.

As usual, we let $\eta_n \in \mathcal{P}(E_n)$ be the normalized distributions given by

$$\eta_n(f) = \gamma_n(f)/\gamma_n(1) \quad \text{and we set} \quad G_{n,\gamma_n} := Q_{n+1,\gamma_n}(1)$$

for any $f \in \mathcal{B}_b(E_n)$. In Section 12.2.1, we shall see that the mass-probability process $(\gamma_n(1), \eta_n)$ satisfies the following measure valued equations

$$\left\{ \begin{array}{rcl} \gamma_{n+1}(1) & = & \eta_n(G_{n,\gamma_n})\ \gamma_n(1) \\ \\ \eta_{n+1} & = & \Psi_{G_{n,\gamma_n}}(\eta_n)\, M_{n+1,\gamma_n} \end{array} \right. \tag{9.39}$$

with the Markov transitions $M_{n,\gamma}$ defined for any $f \in \mathcal{B}_b(E_n)$ and $\gamma \in \mathcal{M}_+(E_{n-1})$ by the following equation

$$M_{n,\gamma}(f) := Q_{n,\gamma}(f)/Q_{n,\gamma}(1)$$

In the above displayed formula, Ψ_G stands for the Boltzmann-Gibbs transformations associated with a given potential function G defined in (7).

To describe with some precision the mean field particle interpretation of these models, we consider a collection of Markov transitions S_{n,η_n} satisfying the compatibility condition

$$\eta_n S_{n,\gamma_n} = \Psi_{G_{n,\gamma_n}}(\eta_n)$$

Several examples of such transitions are provided in (8). In this situation, we can rewrite the evolution equation of the normalized measures as follows:

$$\eta_{n+1} = \eta_n K_{n+1,\gamma_n} \quad \text{with} \quad K_{n+1,\gamma_n} = S_{n,\gamma_n} M_{n+1,\gamma_n}$$

The mean field particle interpretation of the evolution Equations (9.38) is the Markov chain $(\gamma_n^N(1), \xi_n) \in \left(\mathbb{R}_+ \times E_n^N\right)$ defined by

$$\left\{ \begin{array}{rcl} \gamma_{n+1}^N(1) & = & \gamma_n^N(1)\ \eta_n^N(G_{n,\gamma_n^N(1)\eta_n^N}) \\ \\ \mathbb{P}\left(\xi_{n+1} \in dx \ \middle|\ \xi_{n,\gamma_n^N(1)}\right) & = & \displaystyle\prod_{i=1}^{N} K_{n+1,\gamma_n^N(1)\eta_n^N}(\xi_n^i, dx^i) \end{array} \right. \tag{9.40}$$

Chapter 10

Feynman-Kac path integration

10.1 Particle unnormalized measures

We return to the Feynman-Kac models discussed in section 4.2 and in section 9.1.4. We let $K_{n,\eta}$ be some McKean interpretation of the flow of Feynman-Kac measures (9.8). We denote by η_n^N the empirical measures of the mean field N-particle model ξ_n associated with the transitions $K_{n,\eta}$. Mimicking the product formula (4.18), the normalizing constants and the unnormalized distributions (5.37) in the Feynman-Kac model (4.17) can be computed using the following *unbias* estimates

$$\mathcal{Z}_n^N := \prod_{0\leq p<n} \eta_p^N(G_p) \longrightarrow_{N\to\infty} \mathcal{Z}_n = \prod_{0\leq p<n} \eta_p(G_p) \tag{10.1}$$

as well as

$$\gamma_n^N(f_n) := \eta_n^N(f_n) \times \prod_{0\leq p<n} \eta_p^N(G_p) \tag{10.2}$$

Next, we give a simple proof of the unbiasedness properties of these particle models.
By construction, we have

$$\mathbb{E}\left(\gamma_n^N(f_n) \mid \xi_0,\ldots,\xi_{n-1}\right) = \Phi_n\left(\eta_{n-1}^N\right)(f_n) \prod_{0\leq p<n} \eta_p^N(G_p)$$

Notice that

$$\begin{aligned}\Phi_n\left(\eta_{n-1}^N\right)(f_n) &= \Psi_n\left(\eta_{n-1}^N\right)(M_n(f_n)) \\ &= \frac{\eta_{n-1}^N(G_{n-1}M_n(f_n))}{\eta_{n-1}^N(G_{n-1})} = \frac{1}{\eta_{n-1}^N(G_{n-1})}\,\eta_{n-1}^N(Q_n(f_n))\end{aligned}$$

with the integral operator Q_n defined in (4.15) and given by

$$Q_n(f_n) = G_{n-1}M_n(f_n)$$

This shows that

$$\mathbb{E}\left(\gamma_n^N(f_n) \mid \xi_0,\ldots,\xi_{n-1}\right) = \eta_{n-1}^N(Q_n(f_n)) \prod_{0\leq p<(n-1)} \eta_p^N(G_p) = \gamma_{n-1}^N(Q_n(f_n)) \tag{10.3}$$

Iterating the argument, we find that

$$\forall p \leq n \qquad \mathbb{E}\left(\gamma_n^N(f_n) \mid \xi_0,\ldots,\xi_p\right) = \gamma_p^N(Q_{p,n}(f_n)) \Rightarrow \mathbb{E}\left(\gamma_n^N(f_n)\right) = \gamma_n(f_n) \tag{10.4}$$

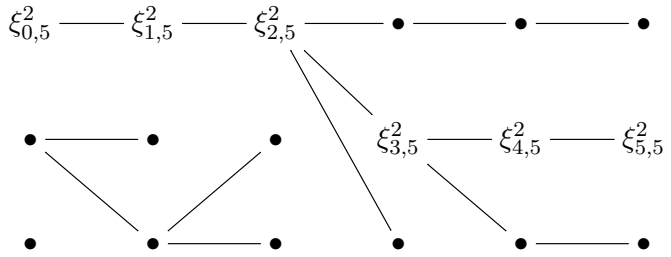
10.2 Genealogical tree occupation measures

The mean field particle interpretation $(\boldsymbol{\xi_n^i})_{1\leq i\leq N}$ of the Feynman-Kac models in path space discussed in (4.16) coincides with the genealogical tree evolution of the particle model discussed above.

More precisely, if we interpret the updating-selection transition as a birth and death process, then arises the important notion of *the ancestral line* of an individual. More precisely, when a particle $\widehat{\xi}_{n-1}^i \longrightarrow \xi_n^i$ evolves to a new location ξ_n^i, we can interpret $\widehat{\xi}_{n-1}^i$ as the parent of ξ_n^i. Looking backwards in time and recalling that the particle $\widehat{\xi}_{n-1}^i$ has selected a site ξ_{n-1}^j in the configuration at time $(n-1)$, we can interpret this site ξ_{n-1}^j as the parent of $\widehat{\xi}_{n-1}^i$ and therefore as the ancestor denoted $\xi_{n-1,n}^i$ at level $(n-1)$ of ξ_n^i. Running backwards in time we may trace the whole ancestral line

$$\xi_{0,n}^i \longleftarrow \xi_{1,n}^i \longleftarrow \ldots \longleftarrow \xi_{n-1,n}^i \longleftarrow \xi_{n,n}^i = \xi_n^i \tag{10.5}$$

An illustration of the genealogical tree model associated with $N = 3$ particles and a time horizon $n = 5$ is given below



Each parent node has a certain random number of children or offspring.

The ancestral line of the i-th individual in the mean field Feynman-Kac population at time n is given by a path-valued particle

$$\boldsymbol{\xi_n^i} := \left(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i\right) \in \boldsymbol{E_n} := (E_0 \times \ldots \times E_n)$$

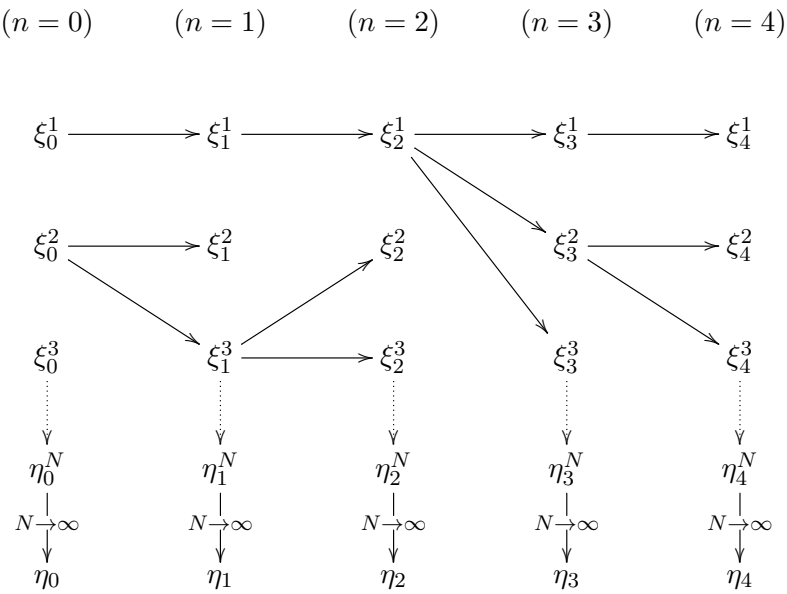
and the occupation measure of the genealogical tree is the empirical measure on path space given by

$$\boldsymbol{\eta_n^N} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\boldsymbol{\xi_n^i}} \longrightarrow_{N\uparrow\infty} \boldsymbol{\eta_n} \tag{10.6}$$

with the Feynman-Kac measures $\boldsymbol{\eta_n} = \mathbb{Q}_n$ in path space discussed in (4.16)

In contrast to the genealogical tree models discussed above, the complete ancestral tree incorporates all the ancestral lineages of the individuals during their evolutions. More formally, the complete ancestral tree model is defined by the set of all the population of individuals $\xi_p = (\xi_p^i)_{1\leq i\leq N}$, from the origin $p = 0$ up to a given final time horizon $p = n$.

An illustration of the complete ancestral tree model associated with $N = 3$ particles, and a time horizon $n = 4$, is given by the following diagram.



10.3 Backward particle measures

Mimicking the backward Markov chain formula (4.20), the measure $\mathbb{Q}_n$ can alternatively be approximated (under the assumption (4.19)) using the backward particle measures

$$\mathbb{Q}_n^N(d(x_0,\ldots,x_n)) = \eta_n^N(dx_n) \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1}) \tag{10.7}$$

with the collection of Markov transitions

$$\begin{aligned}\mathbb{M}_{n+1,\eta_n^N}(x_{n+1}, dx_n) &= \frac{\eta_n^N(dx_n)\ H_{n+1}(x_n, x_{n+1})}{\eta_n^N\left(H_{n+1}(.,x_{n+1})\right)} \\ &= \sum_{1\le i\le N} \frac{H_{n+1}(\xi_n^i, x_{n+1})}{\sum_{1\le j\le N} H_{n+1}(\xi_n^j, x_{n+1})}\ \delta_{\xi_n^i}(dx_n)\end{aligned} \tag{10.8}$$

For any function F_n on $\boldsymbol{E_n}$ we have the unbias property

$$\mathbb{E}\left(\mathbb{Q}_n^N(F_n) \prod_{0\le p<n} \eta_p^N(G_p)\right) = \mathbb{E}\left(F_n(X_0,\ldots,X_n) \prod_{0\le p<n} G_p(X_p)\right) := \Gamma_n(F_n)$$

We set

$$\Gamma_n^N(F_n) := \mathbb{Q}_n^N(F_n) \prod_{0\le p<n} \eta_p^N(G_p) \tag{10.9}$$

Arguing as in the proof of (10.3), we have

$$\begin{aligned}&\mathbb{E}\left(\Gamma_n^N(F_n) \mid \xi_0,\ldots,\xi_{n-1}\right) \\ &= \gamma_n^N(1) \int \Phi_n\left(\eta_{n-1}^N\right)(dx_n) \left\{\prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1})\right\} F_n(x_0,\ldots,x_n) \\ &= \gamma_{n-1}^N(1) \int (\eta_{n-1}^N Q_n)(dx_n)\ \mathbb{M}_{n,\eta_{n-1}^N}(x_n, dx_{n-1}) \left\{\prod_{q=1}^{n-1} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1})\right\} F_n(x_0,\ldots,x_n)\end{aligned}$$

Notice that

$$(\eta_{n-1}^N Q_n)(dx_n) = \eta_{n-1}^N(H_n(.,x_n))\ \lambda_n(dx_n)$$

$$\Rightarrow\ (\eta_{n-1}^N Q_n)(dx_n)\ \mathbb{M}_{n,\eta_{n-1}^N}(x_n, dx_{n-1}) = \eta_{n-1}^N(dx_{n-1}) Q_n(x_{n-1}, dx_n)$$

This implies that

$$\mathbb{E}\left(\Gamma_n^N(F_n) \mid \xi_0, \ldots, \xi_{n-1}\right)$$

$$= \gamma_n^N(1) \int \Phi_n\left(\eta_{n-1}^N\right)(dx_n) \left\{\prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1})\right\} F_n(x_0, \ldots, x_n)$$

$$= \gamma_{n-1}^N(1) \int \eta_{n-1}^N(dx_{n-1}) \left\{\prod_{q=1}^{n-1} \mathbb{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1})\right\} F_{n-1,n}(x_0, \ldots, x_{n-1}) = \Gamma_{n-1}^N(F_{n-1,n})$$

with

$$F_{n-1,n}(x_0, \ldots, x_{n-1}) = \int_{E_n} Q_n(x_{n-1}, dx_n)\ F_n(x_0, \ldots, x_n)$$

Iterating the argument, we prove that

$$\forall p \leq n \qquad \mathbb{E}\left(\Gamma_n^N(F_n) \mid \xi_0, \ldots, \xi_p\right) = \Gamma_p^N(F_{p,n}) \implies \mathbb{E}(\Gamma_n^N(F_n)) = \Gamma_n(F_n)$$

with the collection of functions

$$F_{p,n}(x_0, \ldots, x_p) = \int_{E_{p+1}\times\ldots\times E_n} Q_{p+1}(x_p, dx_{p+1}) \ldots Q_n(x_{n-1}, dx_n)\ F_n(x_0, \ldots, x_n)$$

10.4 A random particle matrix model

The computation of integrals w.r.t. the particle measures $\mathbb{Q}_n^N$ are reduced to summations over the particle locations ξ_n^i. It is therefore natural to identify a population of individuals $(\xi_n^1, \ldots, \xi_n^N)$ at time n to the ordered set of indexes $\{1, \ldots, N\}$. In this framework, the occupation measures and the functions are identified with the following line and column vectors

$$\eta_n^N := \left[\frac{1}{N}, \ldots, \frac{1}{N}\right] \quad \text{and} \quad f_n := \begin{pmatrix} f_n(\xi_n^1) \\ \vdots \\ f_n(\xi_n^N) \end{pmatrix}$$

and the matrices $\mathbb{M}_{n,\eta_{n-1}^N}$ by the $(N \times N)$ matrices

$$\mathbb{M}_{n,\eta_{n-1}^N} := \begin{pmatrix} \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^N) \\ \vdots & \vdots & \vdots \\ \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^N) \end{pmatrix} \tag{10.10}$$

with the (i,j)-entries

$$\mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^i, \xi_{n-1}^j) = \frac{G_{n-1}(\xi_{n-1}^j) H_n(\xi_{n-1}^j, \xi_n^i)}{\sum_{k=1}^N G_{n-1}(\xi_{n-1}^k) H_n(\xi_{n-1}^k, \xi_n^i)}$$

For instance, the $\mathbb{Q}_n$-integration of normalized additive linear functionals of the form

$$\mathbf{f_n}(x_0,\ldots,x_n) = \frac{1}{n+1}\sum_{0\leq p\leq n} f_p(x_p) \tag{10.11}$$

is given by the particle matrix approximation model

$$\mathbb{Q}_n^N(\mathbf{f_n}) = \frac{1}{n+1}\sum_{0\leq p\leq n} \eta_n^N \mathbb{M}_{n,\eta_{n-1}^N}\mathbb{M}_{n-1,\eta_{n-2}^N}\ldots\mathbb{M}_{p+1,\eta_p^N}(f_p)$$

This Markov interpretation allows computing complex Feynman-Kac path integrals using simple random matrix operations on finite sets. Roughly speaking, this methodology allows reducing Feynman-Kac path integration problems on general state spaces to Markov path integration on *finite state spaces*, with cardinality N.

Nevertheless, the computational cost N^2 of these particle random matrix models can be prohibitive in some applications. In this case, we can replace the full matrix averaging technique on the finite sets of N individuals, by some judicious sampling approximation scheme. In this connection, we quote a rejection sampling method, recently proposed in [232].

10.5 Many-body Feynman-Kac models

10.5.1 Feynman-Kac particle models

We write $\mathcal{M}_n$ for the Markov transitions of the particle model $\chi_n{:=}\,\xi_n$ viewed now as a *Markov chain* on $\mathcal{E}_n := E_n^N$, and introduce the potential functions

$$\mathcal{G}_n(\chi_n) = m(\chi_n)(G_n)$$

In the further development of this section we use calligraphic letters such as x_n and $y_n = (y_n^i)_{1\leq i\leq N}$ to denote states in the product spaces $\mathcal{E}_n$, and slanted roman letters such as x_n, y_n, z_n to denote states in E_n. The path sequences in the product spaces $\boldsymbol{\mathcal{E}}_n := \prod_{0\leq p\leq n}\mathcal{E}_p$ and $\boldsymbol{E}_n := \prod_{0\leq p\leq n} E_p$ are denoted by bold letters such as $\boldsymbol{x_n} = (x_p)_{0\leq p\leq n} \in \boldsymbol{\mathcal{S}}_n$ and $\mathbf{x_n} = (\mathrm{x}_p)_{0\leq p\leq n} \in \boldsymbol{S}_n$. Finally, we also denote by $dx_n = d(\mathrm{x}_n^1,\ldots,\mathrm{x}_n^N)$, resp. $d\boldsymbol{x_n} = d(x_0,\ldots,x_n)$, the infinitesimal neighborhoods of a point $x_n = (\mathrm{x}_n^i)_{1\leq i\leq N} \in \mathcal{E}_n$, resp. $\boldsymbol{x_n} = (x_p)_{0\leq p\leq n} \in \boldsymbol{\mathcal{E}}_n = \prod_{0\leq p\leq n}\mathcal{E}_p$.

We let (Π_n,Λ_n) be the Feynman-Kac measures on $\mathcal{E}_n$ defined for any $\mathcal{F}_n \in \mathcal{B}(\mathcal{E}_n)$ by

$$\Pi_n(\mathcal{F}_n) := \Lambda_n(\mathcal{F}_n)/\Gamma_n(1) \quad \text{with} \quad \Lambda_n(F_n) = \mathbb{E}\left(\mathcal{F}_n(\chi_n)\prod_{0\leq p<n}\mathcal{G}_p(\chi_p)\right) \tag{10.12}$$

Notice that the unbiasedness properties of unnormalized particle measures ensures that $\Lambda_n(1) = \gamma_n(1)$. It is also readily checked that

$$\Lambda_{n+1} = \Lambda_n\mathcal{Q}_{n+1} \quad \text{and} \quad \Pi_{n+1} := \Psi_{\mathcal{G}_n}(\Pi_n)\mathcal{M}_{n+1} \tag{10.13}$$

with the integral operators

$$\mathcal{Q}_{n+1}(x_n,dx_{n+1}) = \mathcal{G}_n(x_n)\ \mathcal{M}_{n+1}(x_n,dx_{n+1})$$

We denote by $(\boldsymbol{\Pi_n},\boldsymbol{\Lambda_n})$ the Feynman-Kac measures associated with the historical process

$$\boldsymbol{\chi_n}{=}\,(\chi_0,\ldots,\chi_n)$$

and the potential functions $\boldsymbol{\mathcal{G}_n(\chi_n)} := \mathcal{G}_n(\chi_n)$ on the path space $\boldsymbol{\mathcal{E}}_n = (\mathcal{E}_0 \times \ldots \times \mathcal{E}_n)$. More formally, these measures are defined for any $\boldsymbol{\mathcal{F}_n} \in \mathcal{B}(\boldsymbol{\mathcal{E}_n})$ by

$$\boldsymbol{\Pi_n(\mathcal{F}_n) := \Lambda_n(\mathcal{F}_n)/\Lambda_n(1)} \quad \text{with} \quad \boldsymbol{\Lambda_n(\mathcal{F}_n)} = \mathbb{E}\left(\boldsymbol{\mathcal{F}_n(\chi_n)} \prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p(\chi_p)}\right) \tag{10.14}$$

Whenever the integral operators Q_n have some densities H_n w.r.t. some reference distributions λ_n on E_n, given $\boldsymbol{\chi_n}$ we let $\mathbb{X}_{\mathbf{n}} := (\mathbb{X}_p)_{0\leq p\leq n}$ be a random path with conditional distribution

$$\boldsymbol{\mathcal{K}_n(\chi_n, dx_n)} := m(\chi_n)(dx_n) \prod_{1\leq k\leq n} \mathbb{M}_{k,m(\chi_{k-1})}(x_k, dx_{k-1}) \tag{10.15}$$

In the above displayed formula $\boldsymbol{dx_n}$ stands for an infinitesimal neighborhood of the path $\mathbf{x_n} = (x_p)_{0\leq p\leq n} \in \boldsymbol{E_n}$, and $\mathbb{M}_{k,m(x_{k-1})}$ are the Markov transitions defined in (10.8).

The unbiasedness properties of unnormalized particle measures are equivalent to the fact that for any $(\boldsymbol{f_n}, f_n) \in (\mathcal{B}(\boldsymbol{E_n}) \times \mathcal{B}(E_n))$, we have

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{f_n}(\mathbb{X}_{\boldsymbol{n}}) \prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p(\chi_p)}\right) &= \mathbb{E}\left(\boldsymbol{f_n}(\boldsymbol{X_n}) \prod_{0\leq p<n} \boldsymbol{G_p(X_p)}\right) \\
\mathbb{E}\left(f_n(\mathbb{X}_n) \prod_{0\leq p<n} \mathcal{G}_p(\chi_p)\right) &= \mathbb{E}\left(f_n(X_n) \prod_{0\leq p<n} G_p(X_p)\right)
\end{aligned} \tag{10.16}$$

We emphasize that (10.16) holds true for general Feynman-Kac models (i.e. without any regularity on Q_n). In this setting, (10.16) is satisfied with a r.v. $\mathbb{X}_n$ with conditional distribution given $\boldsymbol{\chi_n}$ defined by

$$\mathcal{K}_n(\boldsymbol{\chi_n}, dx_n) = m(\chi_n)(dx_n) \tag{10.17}$$

The measures (Π_n, Λ_n) and their path space versions $(\boldsymbol{\Pi_n, \Lambda_n})$ are called the many body Feynman-Kac measures associated with the mean field N-particle interpretation of the measures (η_n, γ_n). As the name "many-body" suggests, these Feynman-Kac models encode properly the collective motion under mean field constraints of the system of particles associated to a standard Feynman-Kac particle system.

From an abstract point of view, all of these measures are of course essentially equivalent to the abstract Feynman-Kac model introduced in section 4.2.1.

10.5.2 Conditional particle Markov chains

We return to the Feynman-Kac models and their many-body versions discussed in section 4.2 and in section 10.5. We only consider the mean field particle model (9.37) with $\epsilon = 0$.

We start the section with a pivotal duality formula between the Feynman-Kac integral operators $(Q_n, \mathcal{Q}_n)$.

Lemma 10.5.1 *We have the duality formula between integral operators on $\mathcal{E}_n \times E_n$*

$$\mathcal{Q}_n(x_{n-1}, dx_n)\ m(x_n)(dx_n) \ = \ (m(x_{n-1})Q_n)(dx_n)\ \mathcal{M}_{x_n,n}(x_{n-1}, dx_n) \tag{10.18}$$

and

$$\eta_0^{\otimes N}(dx_0)\ m(x_0)(dx_0) = \eta_0(dx_0)\ \mu_{x_0}(dx_0)$$

with the collection of Markov transitions

$$\mathcal{M}_{x_n,n}(x_{n-1},dx_n)=\frac{1}{N}\left[\sum_{i=0}^{N-1}\Phi_n(m(x_{n-1}))^{\otimes(i)}\otimes\delta_{x_n}\otimes\Phi_n(m(x_{n-1}))^{\otimes(N-i-1)}\right](dx_n)$$

and the distribution

$$\mu_{x_0}:=\frac{1}{N}\sum_{i=0}^{N-1}\left(\eta_0^{\otimes(i)}\otimes\delta_{x_0}\otimes\eta_0^{\otimes(N-i-1)}\right)$$

Proof :
To check (10.18) we use the symmetry properties of the Markov transitions $\mathcal{M}_n$ to check that for any function $H_n\in\mathcal{B}(E_n\times\mathcal{E}_n)$, we have

$$\begin{aligned}
&\int\ \mathcal{Q}_n(x_{n-1},dx_n)\ m(x_n)(dz_n)\ H_n(z_n,x_n)\\
&=\mathcal{G}_{n-1}(x_{n-1})\ \int\ \Phi_n(m(x_{n-1}))^{\otimes N}(dx_n)\ H_n(x_n^1,x_n)\\
&=m(x_{n-1})(G_{n-1})\ \int\ \Phi_n(m(x_{n-1}))(dx_n^1)\ \left[\delta_{x_n^1}\otimes\Phi_n(m(x_{n-1}))^{\otimes(N-1)}\right](dy_n)\ H_n(x_n^1,y_n)
\end{aligned}$$

The end of the proof comes from the fact that

$$m(x_{n-1})(G_{n-1})\ \Phi_n(m(x_{n-1}))(dx_n^1)=(m(x_{n-1})Q_n)(dx_n^1)$$

The proof of the lemma is now completed. ∎

Definition 10.5.2 *Given a random path $X=(X_n)_{n\geq 0}$ we let $\mathcal{X}_n^\star=(\mathcal{X}_n^{\star,i})_{i=1,\ldots,N}\in\mathcal{S}_n$ be the Markov chain with the transitions $\mathcal{M}_{X_n,n}$, and the initial distribution μ_{X_0} introduced in lemma 10.5.1. We denote by $\mathbb{M}_{\boldsymbol{n}}^\star(\mathbf{X_n},\boldsymbol{dx_n})$ the conditional distributions of the random path $\boldsymbol{\mathcal{X}_n^\star}=\left(\mathcal{X}_p^\star\right)_{0\leq p\leq n}$ on $\boldsymbol{\mathcal{E}_n}$. The process $\mathcal{X}_n^\star$ is called the dual mean field model associated with the Feynman-Kac particle model χ_n and the frozen path X.*

The justification of the "duality" terminology between the processes $\mathcal{X}_n^\star$ and χ_n is discussed in the end of the section. The Feynman-Kac measures $(\boldsymbol{\gamma_n},\boldsymbol{\eta_n})$ and their many body version $(\boldsymbol{\Lambda_n},\boldsymbol{\Pi_n})$ are connected by the following duality theorem which can be seen as an extended version of the unbiasedness properties (10.16).

Theorem 10.5.3 *For any $\boldsymbol{F_n}\in\mathcal{B}(\boldsymbol{\mathcal{E}_n})$ by the following equations*

$$\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\chi_n})\prod_{0\leq p<n}\boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)=\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}_n^\star})\prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)\tag{10.19}$$

When the integral operators Q_n have some densities H_n w.r.t. some reference distributions λ_n, for any $\boldsymbol{F_n}\in\mathcal{B}(\boldsymbol{E_n}\times\boldsymbol{\mathcal{E}_n})$ by the duality formula

$$\mathbb{E}\left(\boldsymbol{F_n}(\mathbb{X}_\mathbf{n},\boldsymbol{\chi_n})\prod_{0\leq p<n}\boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)=\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{X_n},\boldsymbol{\mathcal{X}_n^\star})\prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)\tag{10.20}$$

Proof :
The proof of (10.19) is a direct consequence of (10.18). Indeed, using this formula, we find that

$$\begin{aligned}\mathcal{Q}_n(x_{n-1}, dx_n) &= \int [m(x_{n-1})Q_n](dz_n)\ \mathcal{M}_{z_n,n}(x_{n-1}, dx_n) \\ &= \int m(x_{n-1})(dz_{n-1})\ Q_n(z_{n-1}, dz_n)\ \mathcal{M}_{z_n,n}(x_{n-1}, dx_n)\end{aligned}$$

and therefore

$$\begin{aligned}&\mathcal{Q}_{n-1}(x_{n-2}, dx_{n-1})\mathcal{Q}_n(x_{n-1}, dx_n) \\ &= \int m(x_{n-2})(dz_{n-2})\ Q_{n-1}(z_{n-2}, dz_{n-1})\ Q_n(z_{n-1}, dz_n) \\ &\qquad \times \mathcal{M}_{z_{n-1},n-1}(x_{n-2}, dx_{n-1})\mathcal{M}_{z_n,n}(x_{n-1}, dx_n)\end{aligned}$$

Iterating backward in time we prove (10.19). This ends the proof of the first assertion.

The proof of (10.20) is a also direct consequence of (10.18). To check this claim, first we notice that

$$\begin{aligned}(10.18) \quad &\Leftrightarrow \quad m(x_{n-1})(G_{n-1})\mathcal{M}_n(x_{n-1}, dx_n)\ m(x_n)(dx_n) = (m(x_{n-1})Q_n)(dx_n)\ \mathcal{M}_{x_n,n}(x_{n-1}, dx_n) \\ &\Leftrightarrow \quad \mathcal{M}_n(x_{n-1}, dx_n)\ m(x_n)(dx_n) = \Phi_n\left(m(x_{n-1})\right)(dx_n)\ \mathcal{M}_{x_n,n}(x_{n-1}, dx_n)\end{aligned}$$

Using this formula, we find that

$$\begin{aligned}&\boldsymbol{\Lambda_n}(\boldsymbol{dx_n})\ \textstyle\prod_{0\leq p\leq n} m(x_p)(dx_p) \\ &= \left\{\textstyle\prod_{0\leq p<n} m(x_p)(G_p)\right\} \eta_0^{\otimes N}(dx_0)\ m(x_0)(dx_0)\ \left\{\textstyle\prod_{1\leq p\leq n} \mathcal{M}_p(x_{p-1}, dx_p)\ m(x_p)(dx_p)\right\} \\ &= \left\{\textstyle\prod_{0\leq p<n} m(x_p)(G_p)\right\}\ \left\{\eta_0(dx_0)\ \textstyle\prod_{1\leq p\leq n} \Phi_p(m(x_{p-1}))(dx_p)\right\} \mathbb{M}_{\boldsymbol{n}}^{\star}(\mathbf{x_n}, \boldsymbol{dx_n}) \\ &= \left\{\eta_0(dx_0)\ \textstyle\prod_{1\leq p\leq n} m(x_{p-1})(H_p(., x_p))\ \lambda_p(dx_p)\right\} \mathbb{M}_{\boldsymbol{n}}^{\star}(\mathbf{x_n}, \boldsymbol{dx_n})\end{aligned}$$

The last assertion comes from the fact that

$$m(x_{p-1})(G_{p-1})\ \Phi_p(m(x_{p-1}))(d\mathbf{x}_p) = m(x_{p-1})(H_p(., z_p))\ \lambda_p(d\mathbf{x}_p)$$

On the other hand, we have we have

$$\boldsymbol{\mathcal{K}_n}(\boldsymbol{x_n}, \boldsymbol{dx_n}) := m(x_n)(dx_n) \prod_{1\leq p\leq n} \frac{m(x_{p-1})(dx_{p-1})\ H_p(x_{p-1}, x_p)}{m(x_{p-1})(H_p(., x_p))}$$

where $\boldsymbol{dx_n}$ stands for an infinitesimal neighborhood of the path $\mathbf{x_n} = (x_p)_{0\leq p\leq n} \in \boldsymbol{E_n}$. Recalling that

$$Q_p(x_{p-1}, dx_p) = G_p(x_{p-1})\ M_p(x_{p-1}, dx_p) = H_p(x_{p-1}, x_p)\ \lambda_p(dx_p)$$

This implies that

$$\begin{aligned}&\boldsymbol{\Lambda_n}(\boldsymbol{dx_n})\ \boldsymbol{\mathcal{K}_n}(\boldsymbol{x_n}, \boldsymbol{dx_n}) \\ &= \left\{\eta_0(dx_0)\ \textstyle\prod_{1\leq p\leq n} Q_p(x_{p-1}, dx_p)\right\} \mathbb{M}_{\boldsymbol{n}}^{\star}(\mathbf{x_n}, \boldsymbol{dx_n}) = \boldsymbol{\gamma_n}(\boldsymbol{dx_n})\ \mathbb{M}_{\boldsymbol{n}}^{\star}(\mathbf{x_n}, \boldsymbol{dx_n})\end{aligned}$$

The proof of (10.20) is now completed. This ends the proof of the Theorem. ∎

The following Corollary is a direct consequence of (10.16) and (10.20). It provides a interpretation of the conditional distribution of the dual process $\boldsymbol{\mathcal{X}_n}$ w.r.t. a given frozen trajectory as a conditional many body Feynman-Kac model w.r.t. a random path $\mathbb{X}_{\mathbf{n}}$ sampled with the backward distribution (10.15).

Corollary 10.5.4 *For any $\boldsymbol{F_n} \in \mathcal{B}(\boldsymbol{\mathcal{E}_n})$, and for $\boldsymbol{\eta_n}$-almost every path $\mathbf{x_n}$ we have*

$$\mathbb{E}\left(\mathbf{F_n}(\boldsymbol{\mathcal{X}_n^\star}) \mid \boldsymbol{X_n} = \mathbf{x_n}\right) = \frac{\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}_n})\ \prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p}(\boldsymbol{\mathcal{X}_p}) \mid \mathbb{X}_{\mathbf{n}} = \mathbf{x_n}\right)}{\mathbb{E}\left(\prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p}(\boldsymbol{\mathcal{X}_p}) \mid \mathbb{X}_{\mathbf{n}} = \mathbf{x_n}\right)} \tag{10.21}$$

We end this section with an analytic description of the duality formulae (10.19) and (10.20) in terms of the conditional distributions $\mathbb{M}_{\boldsymbol{n}}^\star$ and $\boldsymbol{\mathcal{K}_n}$ introduced in definition 10.5.2 and in (10.15). Using (10.19) we have

$$\forall \mathbf{x_n} \in \boldsymbol{E_n} \qquad \mathbb{M}_{\boldsymbol{n}}^\star(\mathbf{x_n}, .) \ll \boldsymbol{\eta_n}\mathbb{M}_{\boldsymbol{n}}^\star = \boldsymbol{\Pi_n}$$

Thus, we can define the dual operator $\mathbb{M}_{\boldsymbol{n},\boldsymbol{\eta_n}}^\star$ of $\mathbb{M}_{\boldsymbol{n}}^\star$ from $\mathbb{L}_1(\boldsymbol{\eta_n})$ into $\mathbb{L}_1(\boldsymbol{\Pi_n})$ given for any $\boldsymbol{f_n} \in \mathbb{L}_1(\boldsymbol{\eta_n})$ by

$$\mathbb{M}_{\boldsymbol{n},\boldsymbol{\eta_n}}^\star(\boldsymbol{f_n}) = \frac{\mathbf{d}\left(\boldsymbol{\eta}_{\boldsymbol{n},\boldsymbol{f_n}}\mathbb{M}_{\boldsymbol{n}}^\star\right)}{\mathbf{d}\left(\boldsymbol{\eta_n}\mathbb{M}_{\boldsymbol{n}}^\star\right)} = \frac{\mathbf{d}\left(\boldsymbol{\eta}_{\boldsymbol{n},\boldsymbol{f_n}}\mathbb{M}_{\boldsymbol{n}}^\star\right)}{\mathbf{d}\boldsymbol{\Pi_n}} \quad \text{with} \quad \boldsymbol{\eta}_{\boldsymbol{n},\boldsymbol{f_n}}(\boldsymbol{dx_n}) := \boldsymbol{\eta_n}(\boldsymbol{dx_n})\ \boldsymbol{f_n}(\boldsymbol{x_n})$$

In addition, for any conjugate integers $\frac{1}{p} + \frac{1}{q} = 1$, with $1 \leq p, q \leq \infty$, and any pair of functions $(\boldsymbol{f_n}, \boldsymbol{F_n}) \in (\mathbb{L}_p(\boldsymbol{\eta_n}) \times \mathbb{L}_q(\boldsymbol{\Pi_n}))$ we have

$$\boldsymbol{\Pi_n}\left(\ \boldsymbol{F_n}\ \mathbb{M}_{\boldsymbol{n},\boldsymbol{\eta_n}}^\star(\boldsymbol{f_n})\ \right) = \boldsymbol{\eta_n}\left(\ \mathbb{M}_{\boldsymbol{n}}^\star(\boldsymbol{F_n})\ \boldsymbol{f_n}\ \right) \tag{10.22}$$

These constructions shows that formula (10.20) holds true for general models (i.e. even if the integral operators Q_n don't have a density) where $\mathbb{X}_{\mathbf{n}}$ stands for a random path with conditional distribution $\mathbb{M}_{\boldsymbol{n},\boldsymbol{\eta_n}}^\star(\boldsymbol{\mathcal{X}_n}, .)$ given the historical process $\boldsymbol{\mathcal{X}_n}$. In the reverse angle, we have

$$\forall \boldsymbol{x_n} \in \boldsymbol{\mathcal{E}_n} \qquad \boldsymbol{\mathcal{K}_n}(\boldsymbol{x_n}, .) \ll \boldsymbol{\Pi_n}\boldsymbol{\mathcal{K}_n} = \boldsymbol{\eta_n}$$

Thus (10.20) also implies that $\mathbb{M}_{\boldsymbol{n}}^\star$ coincides with the dual operator $\boldsymbol{\mathcal{K}}_{\boldsymbol{n},\boldsymbol{\Pi_n}}^\star$ of $\boldsymbol{\mathcal{K}_n}$ from $\mathbb{L}_1(\boldsymbol{\Pi_n})$ into $\mathbb{L}_1(\boldsymbol{\eta_n})$; that is, we have that

$$(10.20) \implies \boldsymbol{\Pi_n}\boldsymbol{\mathcal{K}_n} = \boldsymbol{\eta_n} \implies \boldsymbol{\eta_n}\left(\boldsymbol{f_n}\ \boldsymbol{\mathcal{K}}_{\boldsymbol{n},\boldsymbol{\Pi_n}}^\star(\boldsymbol{F_n})\right) = \boldsymbol{\Pi_n}\left(\boldsymbol{F_n}\ \boldsymbol{\mathcal{K}_n}(\boldsymbol{f_n})\right)$$

with

$$\boldsymbol{\mathcal{K}}_{\boldsymbol{n},\boldsymbol{\Pi_n}}^\star(\mathbf{z_n}, \boldsymbol{dx_n}) = \boldsymbol{\Pi_n}(\boldsymbol{dx_n})\ \frac{\boldsymbol{d\mathcal{K}_n}(\boldsymbol{x_n}, .)}{\boldsymbol{d\Pi_n}\boldsymbol{\mathcal{K}_n}}(\mathbf{z_n}) = \mathbb{M}_{\boldsymbol{n}}^\star\left(\mathbf{z_n}, \boldsymbol{dx_n}\right) \tag{10.23}$$

These formulations underline the duality between the random paths $\boldsymbol{\mathcal{X}_n^\star}$ and $\mathbb{X}_n$ under the Feynman-Kac measures $\boldsymbol{\eta_n}$ and their many-body version $\boldsymbol{\Pi_n}$.

10.5.3 Historical processes

Let us suppose that $(\eta_n, \gamma_n, \xi_n)$ is the historical version of an auxiliary Feynman-Kac model $(\gamma'_n, \eta'_n, \xi'_n)$ associated with some potential functions G'_n and some Markov chain X'_n transitions M'_n on some state spaces E'_n. In this situation, the reference Markov chain of the Feynman-Kac models (η_n, γ_n) coincides with the historical process $X_n = (X'_0, \ldots, X'_n)$ of the chain X'_n. We also recall that the particle model

$\chi_n := \xi_n$ represents the evolution of the genealogical tree model associated with the particle model $\chi'_n := \xi'_n$.

The same property holds true at the level of the dual processes. More precisely, the dual mean field model $\mathcal{X}^\star_n$ associated with pair (ξ_n, X_n) represents the evolution of the genealogical tree model of the dual particle model $\mathcal{X}^{\star,\prime}_n$ associated with the pair (ξ'_n, X'_n). To be more precise, we observe that the i-th path space particle

$$\mathcal{X}^i_n = \left(\mathcal{X}'^i_{0,n}, \mathcal{X}'^i_{1,n}, \ldots, \mathcal{X}'^i_{n,n}\right) \in E_n := (E'_0 \times \ldots \times E'_n)$$

of the particle model $\mathcal{X}^\star_n$ can be interpreted as the line of ancestors $\mathcal{X}^{\star,\prime i}_{p,n}$ of the i-th individual $\mathcal{X}^{\star,\prime i}_{n,n}$ at time n, at every level $0 \leq p \leq n$, with $1 \leq i \leq N$. This shows that the particle model $\mathcal{X}^{\star,\prime}_n = \left(\mathcal{X}^{\star,\prime i}_n\right)_{1\leq i\leq N}$ coincides with the evolution of the individuals $\mathcal{X}^{\star,\prime}_{n,n} = \left(\mathcal{X}^{\star,\prime i}_{n,n}\right)_{1\leq i\leq N}$.

It is also important to observe that *the dual process $\mathcal{X}_n$ is defined in terms of frozen historical paths* $X_n = (X'_0, \ldots, X'_n)$. Therefore, for any function $\boldsymbol{F_n} \in \mathcal{B}(\boldsymbol{E_n})$, we have the $\boldsymbol{\eta_n}$-almost sure conditional expectation formula

$$\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_n}) \mid \boldsymbol{X_n}\right) = \mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_n}) \mid X_n\right) := \mathbb{M}^\star_n(\boldsymbol{F_n})(X_n) \tag{10.24}$$

In the further development of this section, we denote by $\mathcal{G}'_n$ the potential function of the many-body model associated with the Feynman-Kac model $(\gamma'_n, \eta'_n, \xi'_n)$; that is, we have that $\mathcal{G}'_n(\chi'_n) = m(\chi'_n)(G'_n)$. In this notation, formula (10.19) takes the form

$$\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\chi_n}) \prod_{0\leq p<n} \mathcal{G}'_p(\chi'_p)\right) = \mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_n}) \prod_{0\leq p<n} G'_p(X'_p)\right) \tag{10.25}$$

Choosing a function $\boldsymbol{F_n}$ that only depends on the marginal populations we find that

$$\boldsymbol{F_n}(\chi_0, \ldots, \chi_n) := F_n(\chi'_0, \ldots, \chi'_n)$$

$$\Rightarrow \mathbb{E}\left(F_n(\boldsymbol{\chi'_n}) \prod_{0\leq p<n} \mathcal{G}'_p(\chi'_p)\right) = \mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}'_n}) \prod_{0\leq p<n} G'_p(X'_p)\right)$$

Notice that $\boldsymbol{\chi'_n}$ and $\boldsymbol{\mathcal{X}^{\star,\prime}_n}$ are $\boldsymbol{\mathcal{E}'_n} = \prod_{0\leq p\leq n} \mathcal{E}'_p$ valued random paths with $\mathcal{E}'_n := E'^N_n$, for any $n \geq 0$.

In much the same way, when the integral operators Q'_n have some densities H'_n w.r.t. some reference distributions λ'_n on E'_n, the formula (10.20) takes the following form

$$\mathbb{E}\left(F_n(\mathbb{X}'_n, \boldsymbol{\chi'_n}) \prod_{0\leq p<n} \mathcal{G}'_p(\chi'_p)\right) = \mathbb{E}\left(F_n(X_n, \boldsymbol{\mathcal{X}^{\star,\prime}_n}) \prod_{0\leq p<n} G'_p(X'_p)\right) \tag{10.26}$$

where $\mathbb{X}'_n := (\mathbb{X}'_p)_{0\leq p\leq n}$ stands for a random path in E_n with distribution

$$\boldsymbol{\mathcal{K}'_n}(\boldsymbol{\chi'_n}, dx_n) := m(\chi'_n)(dx'_n) \prod_{1\leq k\leq n} \mathbb{M}_{k,m(\chi'_{k-1})}(x'_k, dx'_{k-1})$$

The following Corollary shows that the transport equations imply an interpretation of mean field particle models with frozen trajectories as conditional many body Feynman-Kac models w.r.t. an random ancestral path $\mathbb{X}_n$ of the process χ'_n.

Corollary 10.5.5 *For any $n \geq 0$, $F_n \in \mathcal{B}(\boldsymbol{\mathcal{E}_n} \times E_n)$ we have*

$$\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\chi_{n-1}},\mathbb{X}_n)\prod_{0\leq p<n}\mathcal{G}'_p(\chi'_p)\right)=\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},X_n)\prod_{0\leq p<n}G'_p(X'_p)\right)\tag{10.27}$$

where $\mathbb{X}_n$ stands for a random path with conditional distribution $m(\chi_n)$, given $\boldsymbol{\chi_n}$. In addition, for any $F_n\in\mathcal{B}(\boldsymbol{\mathcal{E}_n})$ and η_n-almost every $x_n\in E_n$ we have that

$$\frac{\mathbb{E}\left(F_n(\boldsymbol{\chi_{n-1}})\ \prod_{0\leq p<n}\mathcal{G}'_p(\chi'_p)\mid\mathbb{X}_n=x_n\right)}{\mathbb{E}\left(\prod_{0\leq p<n}\mathcal{G}'_p(\chi'_p)\mid\mathbb{X}_n=x_n\right)}=\mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}^\star_{n-1}})\mid X_n=x_n\right)$$

Proof :
Using (10.20), for any function $\boldsymbol{F_n}\in\mathcal{B}(\boldsymbol{\mathcal{E}_{n-1}}\times E_n)$ we check that

$$\begin{aligned}
&\mathbb{E}\left(\int\ m(\chi_n)(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\chi_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)\\
&=\mathbb{E}\left(\int\ \Phi_{n-1}(m(\chi_{n-1}))(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\chi_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)\\
&=\mathbb{E}\left(\int\ \Phi_{n-1}(m(\mathcal{X}^\star_{n-1}))(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)\\
&=\mathbb{E}\left(\int\ m(\mathcal{X}^\star_n)(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\mathbb{E}\left(\int\ m(\mathcal{X}^\star_n)(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)\\
&=\tfrac{1}{N}\ \mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},X_n)\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)\\
&\quad+\left(1-\tfrac{1}{N}\right)\ \mathbb{E}\left(\int\ \Phi_n(m(\mathcal{X}^\star_{n-1}))(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},x_n)\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{X_p})\right)
\end{aligned}$$

This implies that

$$\mathbb{E}\left(\int\ m(\chi_n)(dx_n)\ \boldsymbol{F_n}(\boldsymbol{\chi_{n-1}},x_n)\prod_{0\leq p<n}\boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)=\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\mathcal{X}^\star_{n-1}},X_n)\prod_{0\leq p<n}G_p(X_p)\right)$$

The end of the proof of (10.27) is now clear. ∎

The next result provides a new interpretation of the backward Markov transition $\boldsymbol{\mathcal{K}'_n}$ in terms of the conditional distribution of a genealogical line given the complete ancestral tree.

Corollary 10.5.6 *When the integral operators Q'_n have some densities H'_n w.r.t. some reference distributions v'_n on S'_n, we have*

$$\mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}'_{n-1}},\mathbb{X}_n)\right)=\mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}'_{n-1}},\mathbb{X}'_n)\right)\tag{10.28}$$

with the random paths $\mathbb{X}_n$ and $\mathbb{X}'_n$ on E_n defined in (10.27) and (10.26). In particular, for any $f_n \in \mathcal{B}(E_n)$ this implies that

$$\mathbb{E}\left(m(\chi_n)(f_n) \mid \boldsymbol{\chi'_{n-1}}\right)$$

$$= \int \Phi_n(m(\chi'_{n-1}))(dx'_n) \left\{ \prod_{1\leq k\leq n} \mathbb{M}_{k,m(\chi'_{k-1})}(x'_k, dx'_{k-1}) \right\} f_n(x'_0, \ldots, x'_n)$$

Proof :
Using (10.27) we have

$$\mathbb{E}\left(\int m(\chi_n)(dx_n)\ F_n(\boldsymbol{\chi'_{n-1}}, x_n)\right) = \mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}^{\star,\prime}_{n-1}}, X_n) \prod_{0\leq p<n} (G_p(X_p)/\mathcal{G}'_p(\mathcal{X}^{\star,\prime}_p))\right)$$

On the other hand, using (10.26) we also have that

$$\mathbb{E}\left(F_n(\boldsymbol{\chi'_{n-1}}, \mathbb{X}'_n)\right) = \mathbb{E}\left(F_n(\boldsymbol{\mathcal{X}^{\star,\prime}_{n-1}}, X_n) \prod_{0\leq p<n} (G_p(X_p)/\mathcal{G}'_p(\mathcal{X}^{\star,\prime}_p))\right)$$

This clearly ends the proof of the Corollary. ∎

10.5.4 Genealogy and backward sampling models

Definition 10.5.7 *When the integral operators Q_n have some densities H_n w.r.t. some distributions λ_n, we consider the Markov transition from $\boldsymbol{S_n}$ into itself defined by $\mathbb{K}_{\boldsymbol{n}} := \mathbb{M}^{\star}_{\boldsymbol{n}}\boldsymbol{\mathcal{K}}_n$, with the couple of operators $(\mathbb{M}^{\star}_{\boldsymbol{n}}, \boldsymbol{\mathcal{K}}_n)$ introduced in definition 10.5.2 and in (10.15).*

When (η_n, γ_n) is the historical version of an auxiliary Feynman-Kac model (γ'_n, η'_n), we consider the Markov transition from S_n into itself defined by $\mathbb{K}_n := \mathbb{M}^{\star}_n\mathcal{K}_n$, with the couple of operators $(\mathbb{M}^{\star}_n, \mathcal{K}_n)$ introduced in (10.24) and in (10.17).

Proposition 10.5.8 *The Markov transitions $\mathbb{K}_n$, resp. $\mathbb{K}_{\boldsymbol{n}}$ are reversible w.r.t. the probability measures η_n, resp. $\boldsymbol{\eta_n}$*

Three elementary proofs of these regularity properties can be underlined:

- Using (10.20), for any couple of functions $\boldsymbol{f_1}, \boldsymbol{f_2} \in \mathcal{B}(\boldsymbol{E_n})$ we have

$$\mathbb{E}\left(\boldsymbol{\mathcal{K}_n(f_1)}(\boldsymbol{\chi_n})\ \boldsymbol{\mathcal{K}_n(f_2)}(\boldsymbol{\chi_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)$$

$$= \mathbb{E}\left(\boldsymbol{f_1}(\mathbb{X}_{\mathbf{n}})\ \boldsymbol{\mathcal{K}_n(f_2)}(\boldsymbol{\chi_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{\mathcal{G}_p}(\boldsymbol{\chi_p})\right)$$

$$= \mathbb{E}\left(\boldsymbol{f_1}(\mathbf{X_n})\ \boldsymbol{\mathcal{K}_n(f_2)}(\boldsymbol{\mathcal{X}^{\star}_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{G_p}(\boldsymbol{X_p})\right) \propto \mathbb{E}\left(\boldsymbol{f_1}(\boldsymbol{X_n})\ \mathbb{K}_{\boldsymbol{n}}\boldsymbol{(f_2)}(\boldsymbol{X_n}) \textstyle\prod_{0\leq p<n} \boldsymbol{G_p}(\boldsymbol{X_p})\right)$$

  Recalling that $\mathcal{K}_n(\boldsymbol{x_n}, .)$ and $\mathbb{K}_n(x_n, .)$ are the E_n-marginal of the measures $\boldsymbol{\mathcal{K}_n}(\boldsymbol{x_n}, .)$ and $\mathbb{K}_{\boldsymbol{n}}(\mathbf{x_n}, .)$, (for any $\boldsymbol{\eta}_n$-p.s., trajectory $\mathbf{x_n} = (x_p)_{0\leq p\leq n} \in \boldsymbol{E_n}$), for any $(f_1, f_2) \in \mathcal{B}(E_n)^2$ the above result implies that

$$\mathbb{E}\left(m(\chi_n)(f_1)\ m(\chi_n)(f_2)\ \textstyle\prod_{0\leq p<n} \mathcal{G}'_p(\chi'_p)\right) \propto \mathbb{E}\left(f_1(X_n)\ \mathbb{K}_n(f_2)(X_n) \textstyle\prod_{0\leq p<n} G_p(X_p)\right)$$

  By symmetry arguments the reversibility follows.

- Combining the unbiasedness properties of the unnormalized particle measures with the transport equation (10.19) we have

$$\left(\boldsymbol{\eta_n = \Pi_n \mathcal{K}_n} \quad \text{and} \quad \boldsymbol{\Pi_n = \eta_n \mathbb{M}_n^\star}\right) \Longrightarrow \boldsymbol{\eta_n = \eta_n \mathbb{M}_n^\star \mathcal{K}_n = \eta_n \mathbb{K}_n}$$

  In much the same way, using the unbiasedness properties of the unnormalized particle measures we check that

$$\left(\eta_n = \boldsymbol{\Pi_n} \mathcal{K}_n \quad \text{and} \quad \boldsymbol{\Pi_n} = \eta_n \mathbb{M}_n^\star\right) \Longrightarrow \eta_n = \eta_n \mathbb{M}_n^\star \mathcal{K}_n = \eta_n \mathbb{K}_n$$

- The reversibility of $\mathbb{K}_{\boldsymbol{n}} = \boldsymbol{\mathcal{K}^\star_{n,\Pi_n} \mathcal{K}_n}$ is also a direct consequence of the the duality formula (10.22). Indeed, for any $(\boldsymbol{f_1}, \boldsymbol{f_2}) \in \mathbb{L}_2^2(\boldsymbol{\eta_n})$ we have that

$$(10.23) \Rightarrow \quad \boldsymbol{\eta_n}\left(\ \boldsymbol{f_1}\ \mathbb{K}_{\boldsymbol{n}}(\boldsymbol{f_2})\right) = \boldsymbol{\Pi_n}\left(\ \boldsymbol{\mathcal{K}_n}(\boldsymbol{f_1})\ \boldsymbol{\mathcal{K}_n}(\boldsymbol{f_2})\ \right) = \boldsymbol{\eta_n}\left(\ \mathbb{K}_{\boldsymbol{n}}(\boldsymbol{f_1})\ \boldsymbol{f_2}\ \right) \tag{10.29}$$

Since $\mathbb{K}_n(x_n, .)$ is the S_n-marginal of the measures $\mathbb{K}_{\boldsymbol{n}}(\mathbf{x_n}, .)$, we also have

$$(10.29) \Longrightarrow \ \forall (f_1, f_2) \in \mathbb{L}_2^2(\eta_n) \qquad \eta_n\left(\ \mathbb{K}_n(f_1)\ f_2\ \right) = \eta_n\left(\ f_1\ \mathbb{K}_n(f_2)\ \right)$$

Next, we present an elementary proof of the ergodicity of the couple of conditional PMCMC transitions discussed above.

Proposition 10.5.9 *The measure $\boldsymbol{\eta_n}$ and η_n are the unique invariant measures of the Markov transitions $\mathbb{K}_{\boldsymbol{n}}$ and $\mathbb{K}_n$. In addition, we have the estimates*

$$\beta\left(\mathbb{K}_n\right) \vee \beta\left(\mathbb{K}_{\boldsymbol{n}}\right) \leq 1 - \tau_n \ \left(1 - \frac{1}{N}\right)^{n+1} \tag{10.30}$$

for some

$$\tau_n \geq \prod_{0 \leq p < n} g_p \qquad \textit{with} \qquad g_n := \sup_{x,y}(G_n(x)/G_n(y))$$

The estimates (10.30) are direct consequence of the following rather crude uniform estimate

$$\mathbb{K}_{\boldsymbol{n}}(\boldsymbol{f_n})(\mathbf{x_n}) \geq \tau_n \ \left(1 - \frac{1}{N}\right)^{n+1} \ \boldsymbol{\eta_n(f_n)}$$

for any non negative function $\boldsymbol{f_n}$ on $\boldsymbol{E_n}$, and any path sequence $\mathbf{z_n} = (z_p)_{0 \leq p \leq n}$. These lower bounds are easily checked by using the fact that

$$\tau_n^{-1}\ \mathbb{K}_{\boldsymbol{n}}(\boldsymbol{f_n})(\mathbf{x_n})$$

$$\geq \mathbb{E}\left(\left\{\textstyle\prod_{0 \leq p < n} m(\mathcal{X}_p^\star)(G_p/\eta_p(G_p))\right\}\ \boldsymbol{\mathcal{K}_n(f_n)(\mathcal{X}_n^\star)}\ \middle|\boldsymbol{X_n} = \mathbf{x_n}\right) \geq \left(1 - \tfrac{1}{N}\right)^{n+1}\ \boldsymbol{\eta_n(f_n)}$$

For a more thorough discussion on these particle Markov chain Monte Carlo methodologies, we refer the reader to the recent article [197].

10.6 Structural stability properties

10.6.1 Change of reference process

We consider a Feynman-Kac model on the set of paths $\boldsymbol{E_n} = (E_0 \times \ldots \times E_n)$ defined for any measurable function $\boldsymbol{f_n}$ on $\boldsymbol{E_n}$ by

$$\boldsymbol{\gamma_n(f_n)} = \mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{0 \leq p < n} \boldsymbol{G_p(X_p)}\right) \quad \text{and} \quad \boldsymbol{\eta_n(f_n)} = \boldsymbol{\gamma_n(f_n)/\gamma_n(1)}$$

In the above display, $\boldsymbol{X_n}$ stands for the random paths

$$\boldsymbol{X_n} = (X_0, \ldots, X_n)$$

of some Markov chain X_n evolving in the state spaces E_n. We further assume that $X_n = (X'_n, X'_{n+1})$ is the Markov chain in transition space of an auxiliary Markov chain model X'_n evolving in some state spaces E'_n with Markov transitions

$$M'_n(x'_{n-1}, dx'_n) = \mathbb{P}\left(X'_n \in dx'_n \mid X'_{n-1} = x'_{n-1}\right)$$

and some initial condition η'_0.

We let K'_n be the Markov transitions of some Markov chain Y'_n on E'_n

$$K'_n(x'_{n-1}, dx'_n) = \mathbb{P}\left(Y'_n \in dx'_n \mid Y'_{n-1} = x'_{n-1}\right)$$

with initial condition μ'_0. We further assume that $\eta'_0 \ll \nu'_0$ and for any $n \geq 1$

$$M'_n(x_{n-1}, .) \ll K'_n(x_{n-1}, .)$$

and we set

$$Y_n = (Y'_n, Y'_{n+1}) \quad \text{and} \quad \boldsymbol{Y_n} = (Y_0, \ldots, Y_n)$$

In this situation, we have

$$\mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{0 \leq p < n} \boldsymbol{G_n(X_n)}\right) = \mathbb{E}\left(\boldsymbol{f_n(Y_n)} \prod_{0 \leq p < n} \boldsymbol{H_n(Y_n)}\right)$$

with the initial potential functions

$$\forall \mathbf{x_0} \in \mathbf{E_0} = E_0 = E'_0 \times E'_1 \qquad \boldsymbol{H_0(x_0)} = \boldsymbol{G_0(x_0)} \times \frac{d\eta_0}{d\eta'_0}(x_0)$$

and for any $1 \leq p < n$ and any path sequence

$$\mathbf{x_p} = (x_0, \ldots, x_p) \in \boldsymbol{E_p} := (E_0 \times \ldots \times E_p) \quad \text{with} \quad \forall 0 \leq k \leq p \qquad x_k = (x'_k, x'_{k+1}) \in E_k = (E'_k \times E'_{k+1})$$

the potential functions

$$\boldsymbol{H_p(x_p)} = \boldsymbol{G_p(x_p)} \times \frac{dM'_{p+1}(x'_p, .)}{dK'_{p+1}(x'_p, .)}(x'_{p+1})$$

This change of probability Feynman-Kac formula shows that there exists an infinite number of ways to describe a given Feynman-Kac model in terms of a reference Markov chain model. The choice of the reference Markov chain model depends on the problem at hand. We emphasize that the mean field particle model associated with a given Feynman-Kac model strongly depends on the choice of reference exploration Markov chain model.

10.6.2 Quenched and annealed models

We consider a Markov chain model (Θ_n, X_n) on some state spaces $(\Xi_n \times E_n)$ with Markov transitions of the form

$$\mathbb{P}\left((\Theta_n, X_n) \in d(\theta_n, x_n) \mid (\Theta_{n-1}, X_{n-1}) = (\theta_{n-1}, x_{n-1})\right) = K_n(\theta_{n-1}, d\theta_n)\ M_{\theta_n,n}(x_{n-1}, dx_n) \quad (10.31)$$

and an initial condition of the form $\mu_0(d\theta_0)\eta_{\theta_0}(dx_0)$. By construction, Θ_n is a Markov chain on Ξ_n with transitions K_n and initial distribution μ_0. In addition, given some path sequence $\Theta = \theta = (\theta_n)_{n\geq 0}$, the stochastic process X_n is a Markov chain with transitions $M_{\theta_n,n}$ and initial condition η_{θ_0}.

We consider the Feynman-Kac model on the set of paths

$$(\boldsymbol{\Xi_n} \times \boldsymbol{E_n}) = (\Xi_0 \times \ldots \times \Xi_n) \times (E_0 \times \ldots \times E_n)$$

defined for any measurable function $\boldsymbol{F_n}$ on $(\boldsymbol{\Xi_n} \times \boldsymbol{E_n})$ by

$$\boldsymbol{\Gamma_n(F_n)} = \mathbb{E}\left(\boldsymbol{F_n(\Theta_n, X_n)} \prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)}\right) \quad \text{and} \quad \boldsymbol{\mathbb{Q}_n(F_n) = \Gamma_n(F_n)/\Gamma_n(1)} \quad (10.32)$$

In the above display, $\boldsymbol{\Theta_n}$ and $\boldsymbol{X_n}$ stands for the random paths

$$\boldsymbol{\Theta_n} = (\Theta_0, \ldots, \Theta_n) \quad \text{and} \quad \boldsymbol{X_n} = (X_0, \ldots, X_n)$$

and $\boldsymbol{G_n}$ is a collection of non negative potential functions on $(\boldsymbol{\Xi_n} \times \boldsymbol{E_n})$. We denote by $\boldsymbol{\Gamma_n^X}$, resp. $\boldsymbol{\Gamma_n^\Theta}$ the marginal of $\boldsymbol{\Gamma_n}$ w.r.t the variable $\boldsymbol{X_n}$, resp. $\boldsymbol{\Theta_n}$. We also let $\mathbb{Q}_n^{\boldsymbol{X}}$, resp. $\mathbb{Q}_n^{\boldsymbol{\Theta}}$, the marginal of $\mathbb{Q}_n$ w.r.t the variable $\boldsymbol{X_n}$, resp. $\boldsymbol{\Theta_n}$.

We consider the quenched Feynman-Kac model on the set of paths $\boldsymbol{E_n} = (E_0 \times \ldots \times E_n)$ defined for any measurable function $\boldsymbol{f_n}$ on $\boldsymbol{E_n}$, and for any $\boldsymbol{\Theta_n = \theta_n} \in \boldsymbol{\Xi_n}$ by

$$\boldsymbol{\gamma_{\theta_n,n}(f_n)} = \mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{p\leq q<n} \boldsymbol{G_{\theta_n,n}(X_n)} \mid \boldsymbol{\Theta_n = \theta_n}\right) \quad \text{and} \quad \boldsymbol{\eta_{\theta_n,n}(f_n) = \gamma_{\theta_n,n}(f_n)/\gamma_{\theta_n,n}(1)} \quad (10.33)$$

In the above display, $\boldsymbol{G_{\theta_n,n}}$ stands for the potential functions defined for any $\boldsymbol{\theta_n} \in \boldsymbol{\Xi_n}$ by

$$\boldsymbol{G_{\theta_n,n}(X_n) = G_n(\theta_n, X_n)}$$

We recall that

$$\mathbb{E}\left(\prod_{0\leq p<n} \boldsymbol{G_{\theta_p,p}(X_p)} \mid \boldsymbol{\Theta_n = \theta_n}\right) = \prod_{0\leq p<n} \boldsymbol{\eta_{\theta_p,p}}\left(\boldsymbol{G_{\theta_p,p}}\right) := \prod_{0\leq p<n} \boldsymbol{h_p(\theta_p)}$$

with the potential functions

$$\boldsymbol{h_p(\theta_p)} := \boldsymbol{\eta_{\theta_p,p}}\left(\boldsymbol{G_{\theta_p,p}}\right)$$

When, the potential functions only depend on the terminal state of the historical process;

$$\boldsymbol{G_n(\theta_n, X_n) = G_{\theta_n,n}(X_n)} = G_n(\theta_n, X_n) = G_{\theta_n,n}(X_n) \quad (10.34)$$

the measures (10.32) reduce to

$$\boldsymbol{\Gamma_n(F_n)} = \mathbb{E}\left(\boldsymbol{F_n(\Theta_n, X_n)} \prod_{0\leq p<n} G_p(\Theta_p, X_p)\right) \quad (10.35)$$

but the potential functions $\boldsymbol{h_p(\theta_p)}$ still depend on the full historical process $\boldsymbol{\theta_p} = (\theta_q)_{0\leq q\leq p}$ with the formula

$$\boldsymbol{h_p(\theta_p)} := \eta_{\boldsymbol{\theta_p},p}\left(G_{\theta_p,p}\right)$$

We let $(\boldsymbol{\nu_n}, \boldsymbol{\mu_n})$ be the Feynman-Kac measures on $\boldsymbol{\Xi_n}$ defined for any $F_n \in \mathcal{B}(\boldsymbol{\Xi_n})$ by

$$\boldsymbol{\nu_n(F_n)} = \mathbb{E}\left(\boldsymbol{F_n(\Theta_n)} \prod_{0\leq p<n} \boldsymbol{h_p(\Theta_p)}\right) \quad \text{and} \quad \boldsymbol{\mu_n(F_n)} = \boldsymbol{\nu_n(F_n)/\nu_n(1)} \tag{10.36}$$

For any $F_n \in \mathcal{B}(\boldsymbol{\Xi_n})$ we have that

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{F_n(\Theta_n)} \prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)}\right) &= \mathbb{E}\left(\boldsymbol{F_n(\Theta_n)}\, \mathbb{E}\left(\prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)} \mid \boldsymbol{\Theta_n}\right)\right) \\ &= \mathbb{E}\left(\boldsymbol{F_n(\Theta_n)} \prod_{0\leq p<n} \boldsymbol{h_p(\Theta_p)}\right)\end{aligned}$$

In the same vein, for any $\boldsymbol{f_n} \in \mathcal{B}(\boldsymbol{E_n})$, we have that

$$\begin{aligned}&\mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)}\right) \\ &= \mathbb{E}\left(\frac{\mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)} \mid \boldsymbol{\Theta_n}\right)}{\mathbb{E}\left(\prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)} \mid \boldsymbol{\Theta_n}\right)} \times \mathbb{E}\left(\prod_{0\leq p<n} \boldsymbol{G_p(\Theta_p, X_p)} \mid \boldsymbol{\Theta_n}\right)\right) \\ &= \mathbb{E}\left(\boldsymbol{\eta_{\Theta_n,n}(f_n)} \prod_{0\leq p<n} \boldsymbol{h_p(\Theta_p)}\right) = \mathbb{E}\left(\boldsymbol{F_{f_n}(\Theta_n)} \prod_{0\leq p<n} \boldsymbol{h_p(\Theta_p)}\right)\end{aligned} \tag{10.37}$$

with the function

$$\boldsymbol{F_{f_n}(\theta_n)} := \boldsymbol{\eta_{\theta_n,n}(f_n)}$$

In summary, the annealed measures $\boldsymbol{\Gamma_n^X}$, resp. $\boldsymbol{\Gamma_n^\Theta}$ can be expressed in terms of the $\boldsymbol{\nu_n}$ with the formulae

$$\boldsymbol{\Gamma_n^\Theta} = \boldsymbol{\nu_n} \qquad \mathbb{Q}_n^{\boldsymbol{\Theta}} = \boldsymbol{\mu_n} \quad \text{and} \quad \boldsymbol{\Gamma_n^X(f_n)} = \boldsymbol{\nu_n(F_{f_n})} \qquad \mathbb{Q}_n^{\boldsymbol{X}}(\boldsymbol{f_n}) = \boldsymbol{\mu_n(F_{f_n})} \tag{10.38}$$

We further assume that the potential functions $\boldsymbol{G_{\theta_n,n}}$ only depend on the terminal state of the historical process; that is, we have that

$$\boldsymbol{G_{\theta_n,n}(X_n)} = G_{\boldsymbol{\theta_n},n}(X_n) \tag{10.39}$$

for some functions $G_{\boldsymbol{\theta_n},n}$ on E_n. In this situation, using (10.37) we have

$$\boldsymbol{f_n}(x_0,\ldots,x_n) = f_n(x_n) \Rightarrow \mathbb{E}\left(f_n(X_n) \prod_{0\leq k<n} G_{\boldsymbol{\Theta_k},k}(X_k)\right) = \mathbb{E}\left(F_{f_n}(\boldsymbol{\Theta_n}) \prod_{0\leq p<n} \boldsymbol{h_p(\Theta_p)}\right) \tag{10.40}$$

with the function

$$F_{f_n}(\boldsymbol{\theta_n}) := \eta_{\boldsymbol{\theta_n},n}(f_n)$$

where $\eta_{\boldsymbol{\theta_n},n}$ stands for the n-th time marginal of $\boldsymbol{\eta_{\theta_n,n}}$.

We further assume that Markov transitions $M_{\theta_n,n}$ are absolutely continuous with respect to some measures λ_n on E_n, and for any $(x_{n-1},x_n)\in(E_{n-1}\times E_n)$ we have

$$G_{\boldsymbol{\theta_{n-1}},n-1}(x_{n-1})\ M_{\theta_n,n}(x_{n-1},dx_n)=H_{\boldsymbol{\theta_n},n}(x_{n-1},x_n)\ \lambda_n(dx_n)$$

for some density function $H_{\boldsymbol{\theta_n},n}$. In this situation, by (4.20) we have

$$\mathbb{Q}_{\boldsymbol{\theta_n},n}(d(x_0,\ldots,x_n))=\boldsymbol{\eta_{\theta_n,n}}(d(x_0,\ldots,x_n))=\eta_{\boldsymbol{\theta_n},n}(dx_n)\ \prod_{q=1}^{n}\mathbb{M}_{q,\eta_{\boldsymbol{\theta_{q-1}},q-1}}(x_q,dx_{q-1}) \tag{10.41}$$

with the collection of Markov transitions

$$\mathbb{M}_{n+1,\eta_{\boldsymbol{\theta_n},n}}(x_{n+1},dx_n)\ =\ \frac{\eta_{\boldsymbol{\theta_n},n}(dx_n)\ H_{\boldsymbol{\theta_{n+1}},n+1}(x_n,x_{n+1})}{\eta_{\boldsymbol{\theta_n},n}\left(\boldsymbol{\theta_{n+1}},n+1(.,x_{n+1})\right)} \tag{10.42}$$

In this situation, we conclude that

$$\mathbb{E}\left(\boldsymbol{f_n}(\boldsymbol{X_n})\ \prod_{0\leq p<n}\boldsymbol{G_p}(\boldsymbol{\Theta_p},\boldsymbol{X_p})\right)=\mathbb{E}\left(\boldsymbol{F_{f_n}}(\boldsymbol{\Theta_n})\ \prod_{0\leq p<n}\boldsymbol{h_p}(\boldsymbol{\Theta_p})\right)\Longrightarrow\boldsymbol{\Gamma_n^X}(\boldsymbol{f_n})=\boldsymbol{\nu_n}(\boldsymbol{F_{f_n}^{\#}})$$

with the function

$$\boldsymbol{F_{f_n}^{\#}}(\boldsymbol{\theta_n}):=\mathbb{Q}_{\boldsymbol{\theta_n},n}(\boldsymbol{f_n})$$

10.6.3 Particle quenched and annealed models

Given $\boldsymbol{\Theta_n}$, we let $\boldsymbol{\xi_n}$ be an N-mean field particle interpretation of the conditional Feynman-Kac measures $\boldsymbol{\eta_{\theta_n,n}}$ on $\boldsymbol{E_n^N}$ defined in (10.33). We set

$$\overline{\boldsymbol{\Theta}}_{\boldsymbol{n}}:=(\boldsymbol{\Theta_n},\boldsymbol{\chi}_n)\in\overline{\boldsymbol{\Xi}}_{\boldsymbol{n}}:=\left(\boldsymbol{\Xi_n}\times\prod_{0\leq p\leq n}\boldsymbol{E_p^N}\right)\quad\text{with}\quad\boldsymbol{\chi}_n:=(\boldsymbol{\xi_0},\ldots,\boldsymbol{\xi_n})$$

We let $(\boldsymbol{\overline{\nu}_n},\boldsymbol{\overline{\mu}_n})$ be the Feynman-Kac measures on $\overline{\boldsymbol{\Xi}}_{\boldsymbol{n}}$ defined for any $\overline{F}_n\in\mathcal{B}(\overline{\boldsymbol{\Xi}}_{\boldsymbol{n}})$ by

$$\boldsymbol{\overline{\nu}_n}(\boldsymbol{\overline{F}_n})=\mathbb{E}\left(\boldsymbol{\overline{F}_n}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{n}})\ \prod_{0\leq p<n}\boldsymbol{\overline{h}_p}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{p}})\right)\quad\text{and}\quad\boldsymbol{\overline{\mu}_n}(\boldsymbol{\overline{F}_n})=\boldsymbol{\overline{\nu}_n}(\boldsymbol{\overline{F}_n})/\boldsymbol{\overline{\nu}_n}(\mathbf{1}) \tag{10.43}$$

with

$$\boldsymbol{\overline{h}_p}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{p}})=\boldsymbol{\eta_{\Theta_p,p}^N}\left(\boldsymbol{G_{\Theta_p,p}}\right)=\frac{1}{N}\sum_{1\leq i\leq N}\boldsymbol{G_{\Theta_p,p}}(\boldsymbol{\xi_p^i})$$

Using the unbiasedness property of the unnormalized particle measures (10.2), for any $\boldsymbol{\theta_n}\in\boldsymbol{\Xi_n}$ (so that $\boldsymbol{\theta_p}=(\theta_0,\ldots,\theta_p)$ for any $p\leq n$) and any $\boldsymbol{f_n}\in\mathcal{B}(\boldsymbol{E_n})$, we have

$$\begin{aligned}
&\mathbb{E}\left(\boldsymbol{\eta_{\Theta_n,n}^N}(\boldsymbol{f_n})\ \textstyle\prod_{0\leq p<n}\boldsymbol{\eta_{\Theta_p,p}^N}\left(\boldsymbol{G_{\Theta_p,p}}\right)\mid\boldsymbol{\Theta_n}=\boldsymbol{\theta_n}\ \right)\\
&=\boldsymbol{\eta_{\theta_n,n}}(\boldsymbol{f_n})\ \textstyle\prod_{0\leq p<n}\boldsymbol{\eta_{\theta_p,p}}\left(\boldsymbol{G_{\theta_p,p}}\right)\\
&=\mathbb{E}\left(\boldsymbol{f_n}(\boldsymbol{X_n})\ \textstyle\prod_{0\leq p<n}\boldsymbol{G_{\Theta_p,p}}(\boldsymbol{X_p})\ \mid\boldsymbol{\Theta_n}=\boldsymbol{\theta_n}\ \right)
\end{aligned}$$

In much the same way, for any $\boldsymbol{F_n} \in \mathcal{B}(\boldsymbol{\Xi_n})$ and $\boldsymbol{f_n} \in \mathcal{B}(\boldsymbol{E_n})$ we find that

$$\mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\Theta_n}) \prod_{0\leq p<n} \boldsymbol{G_p}(\boldsymbol{\Theta_p}, \boldsymbol{X_p})\right) = \mathbb{E}\left(\boldsymbol{F_n}(\boldsymbol{\Theta_n}) \prod_{0\leq p<n} \overline{\boldsymbol{h}}_{\boldsymbol{p}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{p}})\right) \tag{10.44}$$

and

$$\mathbb{E}\left(\boldsymbol{f_n}(\boldsymbol{X_n}) \prod_{0\leq p<n} \boldsymbol{G_p}(\boldsymbol{\Theta_p}, \boldsymbol{X_p})\right) = \mathbb{E}\left(\overline{\boldsymbol{F}}_{\boldsymbol{f_n}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{n}}) \prod_{0\leq p<n} \overline{\boldsymbol{h}}_{\boldsymbol{p}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{p}})\right)$$

with the function

$$\overline{\boldsymbol{F}}_{\boldsymbol{f_n}}(\overline{\boldsymbol{\theta}}_{\boldsymbol{n}}) := \boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{\theta_n},\boldsymbol{n}}(\boldsymbol{f_n})$$

In summary, the annealed measures $\boldsymbol{\Gamma}^{\boldsymbol{X}}_{\boldsymbol{n}}$, resp. $\boldsymbol{\Gamma}^{\boldsymbol{\Theta}}_{\boldsymbol{n}}$ can be expressed in terms of the $\overline{\boldsymbol{\nu}}_{\boldsymbol{n}}$ with the formulae

$$\overline{\boldsymbol{F}}_{\boldsymbol{n}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{n}}) = \boldsymbol{F_n}(\boldsymbol{\Theta_n}) \Rightarrow \boldsymbol{\Gamma}^{\boldsymbol{\Theta}}_{\boldsymbol{n}}(\boldsymbol{F_n}) = \overline{\boldsymbol{\nu}}_{\boldsymbol{n}}(\overline{\boldsymbol{F}}_{\boldsymbol{n}}) \quad \text{and} \quad \mathbb{Q}^{\boldsymbol{\Theta}}_{n}(\boldsymbol{F_n}) = \overline{\boldsymbol{\mu}}_{\boldsymbol{n}}(\overline{\boldsymbol{F}}_{\boldsymbol{n}}) \tag{10.45}$$

as well as

$$\overline{\boldsymbol{F}}_{\boldsymbol{f_n}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{n}}) := \boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{\Theta_n},\boldsymbol{n}}(\boldsymbol{f_n}) \Rightarrow \boldsymbol{\Gamma}^{\boldsymbol{X}}_{\boldsymbol{n}}(\boldsymbol{f_n}) = \overline{\boldsymbol{\nu}}_{\boldsymbol{n}}(\overline{\boldsymbol{F}}_{\boldsymbol{f_n}}) \quad \text{and} \quad \mathbb{Q}^{\boldsymbol{X}}(\boldsymbol{f_n}) = \overline{\boldsymbol{\mu}}_{\boldsymbol{n}}(\overline{\boldsymbol{F}}_{\boldsymbol{f_n}})$$

We further assume that the potential functions $\boldsymbol{G_{\theta_n,n}}$ only depend on the terminal state of the historical process; that is, we have that

$$\boldsymbol{G_{\theta_n,n}}(\boldsymbol{X_n}) = G_{\boldsymbol{\theta_n},n}(X_n) \tag{10.46}$$

for some functions $G_{\boldsymbol{\theta_n},n}$ on E_n. In this situation, we recall that the measures (10.32) reduce to (10.35) but the potential functions $\boldsymbol{h_p(\theta_p)}$ still depend on the full historical process $\boldsymbol{\theta_p} = (\theta_q)_{0\leq q\leq p}$ with the formula

$$\boldsymbol{h_p}(\boldsymbol{\theta_p}) := \eta_{\boldsymbol{\theta_p},p}\left(G_{\theta_p,p}\right)$$

Nevertheless, in this case we have

$$\begin{aligned} &\boldsymbol{f_n}(x_0,\ldots,x_n) = f_n(x_n) \\ \Rightarrow\ &\mathbb{E}\left(f_n(X_n)\ \textstyle\prod_{0\leq k<n} G_{\boldsymbol{\Theta_k},k}(X_k)\right) = \mathbb{E}\left(F_{f_n}(\boldsymbol{\Theta_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{h_p}(\boldsymbol{\Theta_p})\right) \\ &= \mathbb{E}\left(\overline{F}_{f_n}(\overline{\Theta}_n)\ \textstyle\prod_{0\leq p<n} \overline{\boldsymbol{h}}_{\boldsymbol{p}}(\overline{\boldsymbol{\Theta}}_{\boldsymbol{p}})\right) \end{aligned} \tag{10.47}$$

with the functions

$$F_{f_n}(\boldsymbol{\Theta_n}) := \eta_{\boldsymbol{\Theta_n},n}(f_n) \quad \text{and} \quad \overline{F}_{f_n}(\overline{\Theta}_n) := \eta^N_{\boldsymbol{\Theta_n},n}(f_n)$$

where $\eta_{\boldsymbol{\theta_n},n}$ stands for the n-th time marginal of $\boldsymbol{\eta_{\theta_n,n}}$.

In much the same way, under the assumptions (10.39) and (4.19) using the unbiasedness property of the unnormalized version of the backward particle measures

$$\mathbb{Q}^N_{\boldsymbol{\theta_n},n}(d(x_0,\ldots,x_n)) = \boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{\theta_n},\boldsymbol{n}}(d(x_0,\ldots,x_n)) = \eta^N_{\boldsymbol{\theta_n},n}(dx_n) \prod_{q=1}^{n} \mathbb{M}_{q,\eta^N_{\boldsymbol{\theta_{q-1}},q-1}}(x_q, dx_{q-1})$$

(with the backward transitions $\mathbb{M}_{q,\eta_{\boldsymbol{\theta_{q-1}},q-1}}$ defined in (10.42)) we have

$$\begin{aligned} &\mathbb{E}\left(\mathbb{Q}^N_{\boldsymbol{\Theta_n},n}(\boldsymbol{f_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{\Theta_p},\boldsymbol{p}}\left(\boldsymbol{G_{\Theta_p,p}}\right) \mid \boldsymbol{\Theta_n} = \boldsymbol{\theta_n}\ \right) \\ &= \mathbb{Q}_{\boldsymbol{\theta_n},n}(\boldsymbol{f_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{\eta}_{\boldsymbol{\theta_p},\boldsymbol{p}}\left(\boldsymbol{G_{\theta_p,p}}\right) \\ &= \mathbb{E}\left(\boldsymbol{f_n}(\boldsymbol{X_n})\ \textstyle\prod_{0\leq p<n} \boldsymbol{G_{\Theta_p,p}}(\boldsymbol{X_p})\ \mid \boldsymbol{\Theta_n} = \boldsymbol{\theta_n}\ \right) \end{aligned}$$

We conclude that

$$\boldsymbol{\Gamma_n^X(f_n) = \overline{\nu}_n(\overline{F}_{f_n})} \quad \text{and} \quad \boldsymbol{\mathbb{Q}^X(f_n) = \overline{\mu}_n(\overline{F}^{\#}_{f_n})} \quad \text{with} \quad \boldsymbol{\overline{F}^{\#}_{f_n}(\overline{\theta}_n) := \mathbb{Q}^N_{\overline{\theta}_n,n}(f_n)}$$

10.6.4 Noncommutative models

We consider a Markov chain X_n taking values in some measurable state spaces E_n. We equip the space $\mathbb{R}^d$ with some norm $\|.\|$, and we let $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ be the unit sphere associated with this norm. We consider a collection of potential functions taking values in the space of matrices

$$G_n \; : \; x \in E_n \mapsto G_n(x) \in \mathbb{R}^{d\times d} \quad \text{such that} \quad \|G_n(x).u\| > 0$$

for any $u \in \mathbb{S}^{d-1}$, and any $x \in E_n$. We let $\mathcal{B}_b(E_n, \mathbb{R}^d)$ be Banach space of all bounded measurable function f_n from E_n into $\mathbb{R}^d$. We consider the d-dimensional vector Feynman-Kac measures γ_n defined for any $f_n \in \mathcal{B}_b(E_n, \mathbb{R}^d)$, and any $u_0 \in \mathbb{S}^{d-1}$ by

$$\gamma_n(f_n).u_0 := \mathbb{E}\left(f_n(X_n).\left[\prod_{0\leq p<n} G_p(X_p)\right].u_0 \right)$$

with the directed product of noncommutative matrices

$$\prod_{0\leq p<n} G_p(X_p) := G_{n-1}(X_{n-1}) \ldots G_1(X_1) G_0(X_0)$$

One natural way to turn these vector measure models into the Feynman-Kac models presented in (9.7) is to consider the random walk on the sphere $\mathbb{S}^{d-1}$ defined by the recursion

$$U_{n+1} := G_n(X_n).U_n / \|G_n(X_n).U_n\|$$

with the initial condition $U_0 = u_0$. In this situation, we have

$$\left\| \left[\prod_{0\leq p<n} G_p(X_p)\right].u_0 \right\| = \prod_{0\leq p<n} \boldsymbol{G_p(X_p)}$$

with the Markov chain $\boldsymbol{X_n}$, and the potential functions $\boldsymbol{G_n}$ defined by

$$\boldsymbol{X_n} := (X_n, U_n) \in \boldsymbol{E_n} := (E_n \times \mathbb{S}^{d-1}) \quad \text{and} \quad \boldsymbol{G_n(X_n)} := \|G_n(X_n).U_n\|$$

In this notation, we readily check that

$$\gamma_n(f_n).u_0 := \mathbb{E}\left(\boldsymbol{f_n(X_n)} \prod_{0\leq p<n} \boldsymbol{G_p(X_p)} \right) := \boldsymbol{\gamma_n(f_n)}$$

with the function $\boldsymbol{f_n} \in \mathcal{B}_b(\boldsymbol{E_n})$ defined by $\boldsymbol{f_n(X_n)} := f_n(X_n).U_n$.

10.7 Feynman-Kac sensitivity measures

This section is dedicated to the design of Feynman-Kac sensitivity measures. The analysis of these quantities arises in a variety of application domains. To name a few, in nonlinear filtering, they are used to estimate filter derivatives, as well as the gradient of log-likelihood functions in hidden Markov chain models [203, 204, 205, 232]. In financial mathematics, they are also used to compute option price type sensitivities [82, 102, 103, 274]. In this context, sensitivity measures allow the traders to

determine how sensitive the values of options are to small changes in the set of parameters on which they depend. These parameters include the initial price of assets, the volatility parameter, or the risk free rates.

In financial mathematics literature, these risk measures are often named "Greeks" mainly because they are denoted by Greek letters: the delta is the first derivative of the option value w.r.t. to the underlying price. This quantity can be computed using the mean field simulation schemes associated with the Feynman-Kac representations of the gradients of Markov semigroups developed in Section 10.7.2.

10.7.1 Parametric models

We let $\theta \in \mathbb{R}^d$ be some parameter that may represent some kinetic type parameters related to the free evolution model or to adaptive potential functions. We also consider a collection of functions $G_{\theta,n}$ that depend on θ.

We assume that the free evolution model $X_n^{(\theta)}$ associated to some value of the parameter θ is given by a one-step probability transition $M_{\theta,n}(x, dx')$ so that

$$Q_{\theta,n}(x, dx') := G_{\theta,n-1}(x)\ M_{\theta,n}(x, dx') = H_{\theta,n}(x, x')\ \lambda_n(dx')$$

for some positive density functions $H_{\theta,n}$ and some reference probability measures λ_n. To simplify the presentation, we assume that the initial distribution $\eta_0 = \lambda_0$.

We also assume that the gradient, and the Hessian of the logarithms of these functions, w.r.t. the parameter θ, are well defined.

We let $(\Gamma_{\theta,n}, \mathbb{Q}_{\theta,n})$ be the Feynman-Kac measure associated with a given value of θ, and defined for any bounded measurable function $\mathbf{f_n}$ on the path space $\mathbf{E_n} = (E_0 \times \ldots \times E_n)$ by $\mathbb{Q}_{\theta,n}(\mathbf{f_n}) = \Gamma_{\theta,n}(\mathbf{f_n})/\Gamma_{\theta,n}(1)$, with

$$\Gamma_{\theta,n}(\mathbf{f_n}) = \mathbb{E}\left(\mathbf{f_n}(X_0^{(\theta)}, \ldots, X_n^{(\theta)}) \prod_{0\leq p<n} G_{\theta,p}\left(X_p^{(\theta)}\right)\right) \tag{10.48}$$

We also denote by $(\gamma_{\theta,n}, \eta_{\theta,n})$, the n-th time marginal measures of $(\Gamma_{\theta,n}, \mathbb{Q}_{\theta,n})$.

We observe that

$$\Gamma_{\theta,n}(\mathbf{f_n}) = \boldsymbol{\lambda_n}\left(\mathbf{f_n}\ \exp\left(\mathbb{L}_{\theta,n}\right)\ \right) \quad \text{with} \quad \boldsymbol{\lambda_n} = \otimes_{0\leq p\leq n}\lambda_p$$

and the additive functional

$$\mathbb{L}_{\theta,n}(x_0, \ldots, x_n) := \sum_{p=1}^{n} \log\left(H_{\theta,p}(x_{p-1}, x_p)\right) \tag{10.49}$$

By using simple derivation calculations, we prove that the first order derivative of the option value w.r.t. θ is given by

$$\nabla\Gamma_{\theta,n}(\mathbf{f_n}) = \Gamma_{\theta,n}(\mathbf{f_n}\Lambda_{\theta,n}) \qquad \text{with} \quad \Lambda_{\theta,n} := \nabla\mathbb{L}_{\theta,n} \tag{10.50}$$

$$\nabla^2\Gamma_{\theta,n}(\mathbf{f_n}) = \Gamma_{\theta,n}\left[\mathbf{f_n}(\Lambda_{\theta,n})\Lambda'_{\theta,n} + \mathbf{f_n}\nabla^2\mathbb{L}_{\theta,n}\right] \tag{10.51}$$

Next, we illustrate the above discussion with the sensitivity to changes in the diffusion coefficient of the stochastic Equation (5.2), with $d = 1$. We consider a Markov chain $X_n^{(\theta)}$ that satisfies the following equation

$$X_n^{(\theta)} - X_{n-1}^{(\theta)} = b\left(X_{n-1}^{(\theta)}\right)\ \Delta + \left[\sigma\left(X_{n-1}^{(\theta)}\right) + \theta\ \sigma'\left(X_{n-1}^{(\theta)}\right)\right]\ \left(W_{t_n} - W_{t_{n-1}}\right)$$

for some σ' s.t. $\sigma + \theta\ \sigma' > 0$ for any $\theta \in [0,1]$. In this situation, we have

$$\begin{aligned} &\partial_\theta \sum_{p=1}^{n} \log\left(H_{\theta,p}(x_{p-1}, x_p)\right) \\ &= \sum_{p=1}^{n} \frac{\sigma'(x_{p-1})}{\sigma(x_{p-1}) + \theta\sigma'(x_{p-1})} \left[\left(\frac{(x_p - x_{p-1}) - b(x_{p-1})\Delta}{(\sigma(x_{p-1}) + \theta\sigma'(x_{p-1}))\sqrt{\Delta}}\right)^2 - 1\right] \end{aligned} \tag{10.52}$$

To consider sensitivity to changes in the drift of the stochastic Equation (5.2), with $d = 1$, we assume that $X_n^{(\theta)}$ satisfies equation

$$X_n^{(\theta)} - X_{n-1}^{(\theta)} = \left[b\left(X_{n-1}^{(\theta)}\right) + \theta b'\left(X_{n-1}^{(\theta)}\right)\right]\ \Delta + \sigma\left(X_{n-1}^{(\theta)}\right)\ \left(W_{t_n} - W_{t_{n-1}}\right)$$

for some function b'. In this situation, we have

$$\begin{aligned} &\partial_\theta \sum_{p=1}^{n} \log\left(H_{\theta,p}(x_{p-1}, x_p)\right) \\ &= \sum_{p=1}^{n} \left[(x_p - x_{p-1}) - \left[b(x_{p-1}) + \theta b'(x_{p-1})\right]\Delta\right] \times b'(x_{p-1})/\sigma^2(x_{p-1}) \end{aligned}$$

Now, suppose that changes in potential energy functions are given by $\log G_n = [V_n + \theta V_n']$, for some nonnegative functions V_n and V_n'. In this situation, we have that

$$\partial_\theta \sum_{0 \le p < n} \log\left(G_{\theta,p}(x_p)\right) = -\sum_{0 \le p < n} V_p'(x_p) \tag{10.53}$$

We illustrate these models with a Feynman-Kac model associated with a reference Markov chain $X_n^{(\theta)} = X_n$, whose values do not depend on θ

$$\gamma_{\theta,n}(f_n) = \mathbb{E}\left(f_n(X_n)\ \exp\left\{-\sum_{0 \le q < n} V_{\theta,q}(X_q)\right\}\right)$$

for some smooth functions $\theta \mapsto V_{\theta,n}$. Then, using the backward Markov chain model we have

$$\nabla\gamma_{\theta,n}(f_n) \quad = \quad -\sum_{0 \le p < n} \gamma_{\theta,n}\left(f_n\ \mathbb{M}_{n,\eta_{\theta,n-1}} \ldots \mathbb{M}_{p+1,\eta_{\theta,p}}\left(\nabla V_{\theta,p}\right)\right)$$

10.7.2 Gradient of semigroups

We consider a Markov chain model given by an iterated $\mathbb{R}^d$-valued random process

$$X_{n+1} := F_n(X_n) = (F_n \circ F_{n-1} \circ \cdots \circ F_0)(X_0) \tag{10.54}$$

starting at some random state X_0, with a sequence of random smooth functions of the form

$$F_n(x) = \mathcal{F}_n(x, W_n) \tag{10.55}$$

In the above display, W_n is a collection of independent r.v. taking values in $\mathbb{R}^{d'}$, for some $d' \ge 1$; and $\mathcal{F}_n$ are some smooth functions, from $\mathbb{R}^{d+d'}$ into $\mathbb{R}^d$.

The semigroup of the Markov chain X_n is the expectation operator defined for any regular function f_n and any state x by

$$P_{n+1}(f_{n+1})(x) := \mathbb{E}\left(f_{n+1}(X_{n+1}) \mid X_0 = x\right) = \mathbb{E}\left(f(X_{n+1}(x))\right)$$

with the random flows $(X_n(x))_{n\geq 0}$ defined for any $n \geq 0$ by the following equation

$$X_{n+1}(x) = F_n(X_n(x)) \quad \text{with the initial condition } X_0(x) = x$$

By construction, for any $1 \leq i, j \leq d$ we have the first variational equation

$$\frac{\partial X_{n+1}^i}{\partial x^j}(x) = \sum_{1\leq k\leq d} \frac{\partial F_n^i}{\partial x^k}(X_n(x)) \ \frac{\partial X_n^k}{\partial x^j}(x) \tag{10.56}$$

On the other hand, we have that

$$\frac{\partial P_{n+1}(f)}{\partial x^j}(x) = \mathbb{E}\left(\sum_{1\leq i\leq d} \frac{\partial f}{\partial x^i}(X_{n+1}(x)) \ \frac{\partial X_{n+1}^i}{\partial x^j}(x)\right) \tag{10.57}$$

We denote by $V_n = (V_n^{(i,j)})_{1\leq i,j\leq d}$ and $A_n = (A_n^{(i,j)})_{1\leq i,j\leq d}$ the random $(d \times d)$-matrices with the i-th line and j-th column entries

$$\begin{aligned} V_n^{(i,j)}(x) &= \frac{\partial X_n^i}{\partial x^j}(x) \\ A_n^{(i,j)}(x) &= \frac{\partial F_n^i}{\partial x^j}(x) = \frac{\partial \mathcal{F}_n^i(.,W_n)}{\partial x^j}(x) := \mathcal{A}_n^{(i,j)}(x,W_n) \end{aligned}$$

We mention that the matrix V_n coincides with the Jacobian matrix $J(X_n)(x) = \frac{\partial X_n}{\partial x}(x)$ of the function $X_n \ : \ x \in \mathbb{R}^d \mapsto X_n(x) \in \mathbb{R}^d$.

In this notation, the Equation (10.56) can be rewritten in terms of the following random matrix formulae

$$V_{n+1}(x) \ = \ A_n(X_n(x)) \ V_n(x) := \prod_{p=0}^{n} A_p(X_p(x)) \tag{10.58}$$

In the above display, $\prod_{p=0}^n A_p$ stands for the noncommutative product of the random matrices A_p, taken in the order A_n, A_{n-1},..., A_0.

In the same way, the Equation (10.57) can be rewritten as

$$\nabla P_{n+1}(f_{n+1})(x) = \mathbb{E}\left(\nabla f_{n+1}(X_{n+1}) \ V_{n+1} \mid X_0 = x\right) \tag{10.59}$$

with $V_{n+1} := \prod_{0\leq p\leq n} A_p(X_p)$. For instance, for one dimensional models of the form

$$X_{n+1} = \mathcal{F}_n(X_n, W_n) = X_n + b\left(X_n\right) \ \Delta + \sigma\left(X_n\right) \ \sqrt{\Delta} \ W_n \ , \tag{10.60}$$

with some $\Delta > 0$, and some sequence of independent and centered Gaussian random variables W_n, it is readily checked that

$$A_n(x) = \mathcal{A}_n(x, W_n) = \left(1 + \frac{\partial b}{\partial x}(x) \ \Delta + \frac{\partial \sigma}{\partial x}(x) \ \sqrt{\Delta} \ W_n\right)$$

and therefore

$$\begin{aligned} V_{n+1}(x) &= \prod_{p=0}^{n} \left(1 + \frac{\partial b}{\partial x}(X_p)\ \Delta + \frac{\partial \sigma}{\partial x}(X_p)\ \sqrt{\Delta}\ W_p \right) \\ &\simeq_{\Delta \downarrow 0} \exp \sum_{0 \leq p \leq n} \left(\frac{\partial b}{\partial x}(X_p)\ \Delta + \frac{\partial \sigma}{\partial x}(X_p)\ \sqrt{\Delta}\ W_p \right) \end{aligned}$$

In this context (10.59) has the same form as the Feynman-Kac models (9.7).

In the multidimensional case, we proceed as follows. We equip the space $\mathbb{R}^d$ with some norm $\|.\|$. We assume that for any state U_0 in the unit sphere $\mathcal{S}^{d-1} \subset \mathbb{R}^d$, we have

$$\|V_{n+1}\ U_0\| > 0$$

In this situation, we have the multiplicative formulae

$$\nabla f_{n+1}(X_{n+1})\ V_{n+1}\ U_0 = [\nabla f_{n+1}(X_{n+1})\ U_{n+1}] \prod_{0 \leq p \leq n} \|A_p(X_p)\ U_p\|$$

with the well defined $\mathcal{S}^{d-1}$-valued Markov chain defined by

$$U_{n+1} = A_n(X_n)U_n / \|A_n(X_n)U_n\|\ (\Longleftrightarrow U_{n+1} = V_{n+1}\ U_0\ /\ \|V_{n+1}\ U_0\|\ )$$

If we choose $U_0 = u_0$, then we obtain the following Feynman-Kac interpretation of the gradient of a semigroup

$$\nabla P_{n+1}(f_{n+1})(x)\ u_0 = \mathbb{E}\left(F_{n+1}(\mathcal{X}_{n+1}) \prod_{0 \leq p \leq n} \mathcal{G}_p\left(\mathcal{X}_p\right) \right) \tag{10.61}$$

In the above display, $\mathcal{X}_n$ is the Markov chain sequence $\mathcal{X}_n := (X_n, U_n, W_n)$, starting at (x, u_0, W_0); and the functions F_{n+1} and $\mathcal{G}_n$ are defined by

$$F_{n+1}(x, u, w) := \nabla f_{n+1}(x)\ u \quad \text{and} \quad \mathcal{G}_n\left(x, u, w\right) := \|\mathcal{A}_n(x, w)\ u\|$$

We quote the interpolation formula

$$\begin{aligned} P_{n+1}(f)(y) - P_{n+1}(f)(x) &= \int_0^1 \nabla P_{n+1}(f_{n+1})(ty + (1-t)x)\ (y-x)'\ dt \\ &= \mathbb{E}\left(\nabla f(X_{n+1}^{(x,y)}) \left[\prod_{0 \leq p \leq n} A_p(X_p^{(x,y)}) \right] (y-x)' \right) \end{aligned}$$

where $X_n^{(x,y)}$ stands for Markov chain (10.54) starting at some random state $X_0^{(x,y)}$ uniformly chosen in the line segment between x and y.

We end this section with some rather crude upper bound that can be estimated, uniformly w.r.t. the time parameter, under appropriate regularity conditions on the reduced Markov chain model (X_n, W_n). To this end, firstly we notice that

$$\begin{aligned} \mathcal{G}_n\left(x, u, w\right) := \|\mathcal{A}_n(x, w)\ u\| \leq\ G_n(x, w) &:= \|\mathcal{A}_n(x, w)\| \\ &:= \sup_{u \in \mathcal{S}^{d-1}} \|\mathcal{A}_n(x, w)\ u\| \end{aligned}$$

This implies that

$$\begin{aligned} \|\nabla P_{n+1}(f_{n+1})(x)\| &:= \sup_{1 \leq i \leq d} \left| \frac{\partial}{\partial x^i}\ P_{n+1}(f_{n+1})(x) \right| \\ &\leq \|F_{n+1}\| \times \mathbb{E}\left(\prod_{0 \leq p \leq n} G_p\left(X_p, W_p\right) \right) \end{aligned}$$

We end this section with a Feynman-Kac model associated with extremal Lyapunov trajectories. Firstly, in terms of the Jacobian matrices $J_n(X_n)$, we notice that

$$\prod_{0\leq p\leq n} \mathcal{G}_p\left(\mathcal{X}_p\right) = \prod_{0\leq p\leq n} \|A_p(X_p)\ U_p\| = \prod_{0\leq p\leq n} \frac{\|J(X_{p+1})\ u_0\|}{\|J(X_p)\ u_0\|} = \|J(X_{n+1})\ u_0\|$$

Replacing vector norms $\|J(X_n)\ u_0\|$ by matrix norms $\|J(X_n)\|$, and setting

$$\boldsymbol{X_n} = (X_n, X_{n+1}) \quad \text{and} \quad \boldsymbol{G_n(X_n)} = \|\mathrm{Jac}\,(X_{n+1})\|^{\alpha}/\|\mathrm{Jac}\,(X_n)\|^{\alpha}$$

for some $\alpha \in \mathbb{R}$, the Feynman-Kac model described above reduces to the Lyapunov weighted dynamics model

$$\boldsymbol{dQ_n} = \frac{1}{\boldsymbol{\mathcal{Z}_n}}\ \|\mathrm{Jac}\,(X_n)\|^{\alpha}\ \boldsymbol{dP_n}$$

where $\boldsymbol{P_n}$ stands for the law of the random trajectories $(\boldsymbol{X_0}, \ldots, \boldsymbol{X_n})$. When $\alpha < 0$, the measure $\boldsymbol{Q_n}$ favors low Lyapunov trajectories, while for $\alpha > 0$, the $\boldsymbol{Q_n}$ favors high Lyapunov trajectories. This Feynman-Kac model coincides with the rare event stochastic models developed in [180], and presented in Section 11.4.1. The mean field particle models associated with these Feynman-Kac formula have been used in the series of articles [394, 561, 562, 567] to sample atypical rare event trajectories in nonlinear stochastic processes.

We notice that all these Feynman-Kac models we have discussed have exactly the same form as the matrix models discussed in Section 10.6.4. In addition, using the same analysis as above, we easily design mean field estimates of $\mathbb{E}\left(\|\mathcal{Y}_n\|^q\right)$, for any reasonably large $q \geq 0$, just replacing the potential functions $\|Au\|$ by the functions $\|Au\|^q$. We also mention that in physics literature, the mean field IPS simulation models are sometimes called "Resampled Monte Carlo methods" [578].

10.7.3 Malliavin derivatives

We return to the one dimensional model (10.60). For nonsmooth functions f_{n+1} we can use the following Gaussian regularization kernel

$$P_{n+1,\epsilon}(f_{n+1})(x) := \mathbb{E}\left(f_{n+1}(X_{n+1}(x) + \epsilon Y)\right) \tag{10.62}$$

for some auxiliary Gaussian variable, independent of the process X_n. From the statistical viewpoint, this approximation procedure is interpreted as a Gaussian kernel density estimation of the distribution of $X_{n+1}(x)$. Combining (10.59) with (10.62), we end up with the following approximation formula

$$\begin{aligned}&\tfrac{\partial}{\partial x} P_{n+1,\epsilon}(f_{n+1})(x)\\ &= \mathbb{E}\left(\epsilon^{-1}\left[f_{n+1}(X_{n+1}(x) + \epsilon Y) - f_{n+1}(X_{n+1}(x))\right]\ Y\ V_{n+1}(x)\right)\end{aligned}$$

From a more probabilistic point of view, the Gaussian regularization formula (10.62) can also be interpreted as the addition of an additional Gaussian move in the evolution of the chain $X_{n+1}(x)$. This suggests that we can alternatively use the last transition to regularize the semigroup

$$\begin{aligned}P_{n+1}(f_{n+1})(x) &= \mathbb{E}\left(\mathbb{E}\Big(f_{n+1}\big(X_{n+1}(x)\big)\Big|X_n(x)\Big)\right)\\ &= \mathbb{E}\left[\int f_{n+1}\big(x_{n+1}\big)\ \mathbb{P}\left(X_{n+1}(x) \in dx_{n+1}\,|X_n(x)\right)\right]\end{aligned}$$

Letting $H_{n+1}(x_n, x_{n+1})$ be the density of the Markov transition $X_n \rightsquigarrow X_{n+1}$ w.r.t. the Lebesgue measure, we find that

$$P_{n+1}(f_{n+1})(x) = \mathbb{E}\left[\int f_{n+1}\big(x_{n+1}\big)\ H_{n+1}(X_n(x), x_{n+1})\ dx_{n+1}\right]$$

Arguing as above we find that

$$\frac{\partial}{\partial x} P_{n+1}(f_{n+1})(x) = \mathbb{E}\left(f_{n+1}(X_{n+1}(x))\ dH_{n+1}(X_n(x), X_{n+1}(x))\ V_n(x)\right)$$

with the weight function

$$\begin{aligned} dH_{n+1}(x_n, x_{n+1}) &= \frac{\partial}{\partial x_n} \log H_{n+1}(x_n, x_{n+1}) \\ &= \left(\left(\frac{(x_{n+1} - x_n) - b(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}}\right)^2 - 1\right) \frac{\partial}{\partial x} \log \sigma(x_n) \\ &\quad + \left(\frac{(x_{n+1} - x_n) - b(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}}\right) \frac{1 + \frac{\partial b}{\partial x}(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}} \end{aligned}$$

In the context of financial mathematics, these formulae, and the corresponding weighted conventional Monte Carlo approximations, have been recently proposed by N. Chen and P. Glasserman [139]. This framework is an alternative to the Malliavin Greeks derivative calculus introduced in the pioneering articles by E. Fournié, J.M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi [274, 275].

In this connection, we briefly recall some foundations of Malliavin derivatives. We let $P_{s,t}$ be the semigroup associated with the diffusion stochastic equation (5.1), with $d = 1$; that is, we have that

$$P_{s,t}(f)(X_s) = \mathbb{E}\left(f(X_t) \mid X_s\right)$$

Using elementary backward calculations, for any $0 \leq s \leq t$ we find that

$$P_{s,t}(f)(X_s) = P_{0,t}(f)(X_0) + \int_0^s \frac{\partial P_{r,t}(f)}{\partial x}(X_r)\ \sigma(X_r)\ dW_r$$

If we set $s = t$ in the above equation, then we find that

$$\begin{aligned} &\mathbb{E}\left[f(X_t(x))\ \int_0^t \frac{\partial X_s}{\partial x}(x)\ \sigma^{-1}(X_s(x)))\ dW_s\right] \\ &= \mathbb{E}\left[\int_0^t \frac{\partial P_{s,t}(f)}{\partial x}(X_s(x))\ \frac{\partial X_s}{\partial x}(x)\ ds\right] \end{aligned} \tag{10.63}$$

as soon as σ is a regular positive function. Recalling that

$$\frac{\partial}{\partial x} P_{0,t}(f)(x) = \frac{\partial}{\partial x}\mathbb{E}\left[P_{s,t}(f)(X_s(x))\right] = \mathbb{E}\left[\frac{\partial P_{s,t}(f)}{\partial x}(X_s(x))\ \frac{\partial X_s}{\partial x}(x)\right]$$

and using (10.63) we find that

$$\mathbb{E}\left[f(X_t(x))\ \int_0^t \frac{\partial X_s}{\partial x}(x)\ \sigma^{-1}(X_s(x)))\ dW_s\right] = t\ \frac{\partial}{\partial x} P_{0,t}(f)(x)$$

This yields the Malliavin formulation of the semigroup derivatives

$$\frac{\partial}{\partial x} P_{0,t}(f)(x) = \mathbb{E}\left[f(X_t(x))\ \frac{1}{t}\int_0^t \sigma^{-1}(X_s(x))\ \frac{\partial X_s}{\partial x}(x)\ dW_s\right]$$

A more rigorous derivation of the above equations is provided in [274, 275].

Using an Euler type time discretization model

$$X_{(n+1)\Delta} - X_{n\Delta} = b\left(X_{n\Delta}\right)\ \Delta + \sigma\left(X_{n\Delta}\right)\ \sqrt{\Delta}\ Y_n \tag{10.64}$$

with a sequence of independent and centered Gaussian random variables Y_n, we have the Feynman-Kac approximation model

$$\frac{\partial}{\partial x}P_{0,(n+1)\Delta}(f)(x) \simeq_{\Delta\downarrow 0} \frac{1}{(n+1)\sqrt{\Delta}} \sum_{0\leq p\leq n} \mathbb{E}\left(f(X_{(n+1)\Delta}(x))\mathcal{Z}_p(x)\right) \tag{10.65}$$

with the random weight function

$$\mathcal{Z}_p(x) := \varphi\left(X_{p\Delta}(x), Y_p\right) \prod_{0\leq q<p} G_q(X_{q\Delta}(x), Y_q)$$

$$\varphi\left(x,y\right) = \sigma^{-1}(x)\ y \qquad \text{and} \quad G_q(x,y) = 1 + \frac{\partial b}{\partial x}(x)\ \Delta + \frac{\partial \sigma}{\partial x}(x)\ \sqrt{\Delta}\ y$$

The ratio $1/\sqrt{\Delta}$ in the r.h.s. of (10.65) may induce degenerative numerical estimates. One way *to remove* this term in the numerical scheme is to use the following formula

$$\mathbb{E}\left(f(X_{(n+1)\Delta}(x))\mathcal{Z}_p(x)\right) = \mathbb{E}\left(\Upsilon_{p+1,n+1}(f)\left[X_{p\Delta}(x), Y_p\right] \times \mathcal{Z}_p(x)\right)$$

with the function

$$\begin{aligned}\Upsilon_{p+1,n+1}(f)[x,y] \quad &= \quad P_{(p+1)\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta + \sigma(x)\sqrt{\Delta}y\right) \\ &\qquad -P_{(p+1)\Delta,(n+1)\Delta}\left(f\right)\left(x + b\left(x\right)\Delta\right)\end{aligned}$$

Under some appropriate regularity conditions, we notice that

$$\begin{aligned}&\Upsilon_{p+1,n+1}(f)[x,y] \\ &\simeq_{\Delta\downarrow 0}\ P_{p\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta + \sigma(x)\sqrt{\Delta}y\right) - P_{p\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta\right) \\ &\simeq_{\Delta\downarrow 0}\ \frac{\partial P_{p\Delta,(n+1)\Delta}(f)}{\partial x}\left(x\right)\ \sigma(x)\sqrt{\Delta}\ y\end{aligned}$$

This implies that the gradient of the Markov semigroup is approximated by

$$\frac{\partial}{\partial x}P_{0,(n+1)\Delta}(f)(x) \simeq_{\Delta\downarrow 0} \frac{1}{(n+1)} \sum_{0\leq p\leq n} U_{p,n}(f)(x)$$

with the Feynman-Kac formulae

$$U_{p,n}(f)(x) \quad := \quad \mathbb{E}\left(\frac{\partial P_{p\Delta,(n+1)\Delta}(f)}{\partial x}\left(X_{p\Delta}(x)\right)\ Y_p^2 \prod_{0\leq q<p} G_q(X_{q\Delta}(x), Y_q)\right)$$

Chapter 11

Some illustrations

11.1 Ground states of quantum systems

This section is concerned with applications of nonlinear Monte Carlo methods in computational physics. We return to the Feynman-Kac models discussed in section 8.5.2.

11.1.1 Spectral decompositions

We further assume that the infinitesimal generator L of the reference process X_t in (8.48) is a self adjoint operator on $\mathbb{L}_2(\mathbb{R}^d)$ (equipped with the scalar product $\langle f, g\rangle = \int f(x)g(x)dx$) defined on some proper domain of functions $D(L)$; that is, we have that

$$\langle f, L(g)\rangle = \langle L(f), g\rangle$$

for any $f, g \in D(L)$. In this situation, the Schrödinger operator L^V sometimes written in terms of the Hamiltonian operator

$$\mathcal{H} := -L^V = -L + V$$

is again a self adjoint operator on $\mathbb{L}_2(\mathbb{R}^d)$ (under appropriate regularity conditions on V).

An important consequence of the self-adjoint property of $\mathcal{H}$ is that there exists a sequence of non negative eigenvalues $0 \leq E_0 \leq E_1 \leq \ldots$ and a corresponding set of orthonormal eigenfunctions φ_i, $i \geq 0$ (with $\varphi_0 \geq 0$) such that the integral Feynman-Kac operator $Q_t = e^{-t\mathcal{H}}$ introduced in (8.48) and (8.53) has the following spectral representation

$$Q_t(x, dy) = \sum_{i\geq 0} e^{-tE_i}\ \varphi_i(x)\varphi_i(y)\ dy$$

Therefore, we find that

$$Q_t(f)(x) \quad = \quad \sum_{i\geq 0} e^{-tE_i}\ \langle f, \varphi_i\rangle\ \varphi_i(x)$$

By (8.51), we also have that

$$\begin{aligned}\frac{\partial}{\partial t}Q_t(f) &= -\sum_{i\geq 0} E_i\ e^{-tE_i}\ \langle f, \varphi_i\rangle\ \varphi_i \\ &= \sum_{i\geq 0} e^{-tE_i}\ \langle f, \varphi_i\rangle\ L^V(\varphi_i) = L^V(Q_t(f))\end{aligned}$$

Choosing $f = \varphi_i$, we conclude that for any $i \geq 0$

$$\mathcal{H}(\varphi_i) = E_i\ \varphi \Leftrightarrow L^V(\varphi_i) = -E_i\ \varphi_i\ \Rightarrow\ \langle \varphi_i, L^V(\varphi_i)\rangle = -E_i$$

For simplicity, we further assume that $E_0 < E_1$. In this case we have

$$Q_t(f) \simeq_{t\uparrow\infty}\ e^{-tE_0}\ \langle f, \varphi_0\rangle\ \varphi_0$$

This implies that for any starting point x we have

$$-\frac{1}{t}\log Q_t(1)(x)\ \longrightarrow_{t\uparrow\infty}\ E_0 \quad \text{and} \quad \frac{Q_t(f)(x)}{Q_t(1)(x)} \simeq_{t\uparrow\infty}\ \frac{\langle f, \varphi_0\rangle}{\langle 1, \varphi_0\rangle} \tag{11.1}$$

Notice that

$$\frac{\sum_{i\geq 0}\ e^{-tE_i}\ \langle f, \varphi_i\rangle\ \varphi_i(x)}{\sum_{i\geq 0}\ e^{-tE_i}\ \langle 1, \varphi_i\rangle\ \varphi_i(x)} = \frac{\langle f, \varphi_0\rangle\ \varphi_0(x) + \sum_{i\geq 1}\ e^{-t(E_i-E_0)}\ \langle f, \varphi_i\rangle\ \varphi_i(x)}{\langle 1, \varphi_0\rangle\ \varphi_0(x) \sum_{i\geq 1}\ e^{-t(E_i-E_0)}\ \langle 1, \varphi_i\rangle\ \varphi_i(x)}$$

This implies that

$$\frac{Q_t(f)(x)}{Q_t(1)(x)} - \frac{\langle f, \varphi_0\rangle}{\langle 1, \varphi_0\rangle}$$

$$= e^{-t(E_1-E_0)} \sum_{i\geq 1}\ e^{-t(E_i-E_1)}\ \frac{\langle 1, \varphi_i\rangle\ \varphi_i(x)}{\langle 1, \varphi_0\rangle\ \varphi_0(x) +\ \sum_{i\geq 1}\ e^{-t(E_i-E_0)}\ \langle 1, \varphi_i\rangle\ \varphi_i(x)}\ \left[\frac{\langle f, \varphi_i\rangle}{\langle 1, \varphi_i\rangle} - \frac{\langle f, \varphi_0\rangle}{\langle 1, \varphi_0\rangle}\right]$$

We conclude that

$$\frac{Q_t(f)(x)}{Q_t(1)(x)} - \frac{\langle f, \varphi_0\rangle}{\langle 1, \varphi_0\rangle} = \mathrm{O}\left(e^{-t(E_1-E_0)}\right)$$

11.1.2 The harmonic oscillator

The harmonic oscillator is defined by choosing the quadratic energy function

$$\begin{array}{rcl} V(x) = k\ x^2/2 \Rightarrow\ L^V & = & \frac{\hbar^2}{2m}\ \frac{\partial^2}{\partial x} - k\ x^2/2 \\ & = & \frac{\hbar^2}{2m}\ \frac{\partial^2}{\partial x} - \frac{1}{2}\ m\omega^2\ x^2 \quad \text{with} \quad \omega = \sqrt{\frac{k}{m}} \end{array}$$

Whenever they exist, we let φ_n be an eigenfunction of L^V associated with the eigenvalue

$$E_n = \hbar\ \left(n + \frac{1}{2}\right)\ \omega \quad \Longleftrightarrow \quad \frac{2mE_n}{\hbar^2}\ \left(\frac{\hbar^2}{mk}\right)^{1/2} = \frac{2E_n}{\hbar\omega} \quad = (2n+1)$$

$$L^V(\varphi_n) = -E_n\varphi_n \Longleftrightarrow\ \varphi_n''(x) = \left(\frac{mk}{\hbar^2}\ x^2 - \frac{2mE_n}{\hbar^2}\right)\varphi_n(x)$$

Notice that

$$\left(\frac{mk}{\hbar^2}\right)^{1/4} = \frac{1}{\sqrt{\hbar}}\ \left(m\ \left(\frac{k}{m}\right)^{1/2}\right)^{1/2} = \sqrt{\frac{m\omega}{\hbar}}$$

We set

$$\psi_n(x) := \varphi_n\left(\left(\frac{mk}{\hbar^2}\right)^{-1/4} x\right) \iff \varphi_n(x) = \psi_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right)$$

We have

$$\begin{aligned}
\psi_n'(x) &= \left(\frac{mk}{\hbar^2}\right)^{-1/4} \varphi_n'\left(\left(\frac{mk}{\hbar^2}\right)^{-1/4} x\right) \\
\psi_n''(x) &= \left(\frac{mk}{\hbar^2}\right)^{-1/2} \varphi_n''\left(\left(\frac{mk}{\hbar^2}\right)^{-1/4} x\right) \\
&= \left(\frac{\hbar^2}{mk}\right)^{1/2} \left(\frac{mk}{\hbar^2}\left(\left(\frac{mk}{\hbar^2}\right)^{-1/4} x\right)^2 - \frac{2mE_n}{\hbar^2}\right) \psi_n(x)
\end{aligned}$$

This shows that

$$\psi_n''(x) = (x^2 - (2n+1))\ \psi_n(x)$$

Our next objective is to express these eigenfunctions in terms of the Hermite polynomials.

We recall that the Hermite polynomials can be defined using the Rodrigues' formula

$$\mathbb{H}_n(x) = (-1)^n\ e^{x^2}\ \frac{d^n}{dx^n} e^{-x^2}$$

Notice that

$$\begin{aligned}
\mathbb{H}_n'(x) &:= \frac{d\mathbb{H}_n}{dx}(x) = 2x(-1)^n\ e^{x^2}\ \frac{d^n}{dx^n}e^{-x^2} + (-1)^n\ e^{x^2}\ \frac{d^{n+1}}{dx^{n+1}}e^{-x^2} \\
&= 2x\mathbb{H}_n(x) - \mathbb{H}_{n+1}(x) \Leftrightarrow \mathbb{H}_{n+1}(x) = 2x\mathbb{H}_n(x) - \mathbb{H}_n'(x) \qquad (11.2)
\end{aligned}$$

This formula shows that $\mathbb{H}_n$is a polynomial of degree n with a leading coefficient 2^n so that $\frac{d^n\mathbb{H}_n}{dx^n}(x) = 2^n\ n!$. In addition, combining (3.24) with an integration by part we find that

$$\forall m < n \quad \int e^{-x^2}\mathbb{H}_m(x)\mathbb{H}_n(x)\ dx = (-1)^n \int \mathbb{H}_m(x)\ \frac{d^n}{dx^n}e^{-x^2}dx = \int \underbrace{\frac{d^n}{dx^n}\mathbb{H}_m(x)}_{=0}\ e^{-x^2}dx = 0$$

and for $m = n$

$$\int e^{-x^2}\mathbb{H}_n^2(x)\ dx = \int \underbrace{\frac{d^n}{dx^n}\mathbb{H}_n(x)}_{=2^n n!}\ e^{-x^2}dx = 2^n n! \int e^{-x^2}dx = 2^n n!\sqrt{\pi}$$

Working a little more, one deduces that the set of functions

$$\widetilde{\mathbb{H}}_n(x) := (2^n n!\sqrt{\pi})^{-1/2}\ e^{-x^2/2}\mathbb{H}_n(x)$$

forms an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$.

We recall the Leibniz' formula

$$\frac{d^n}{dx^n}(fg) := (fg)^{(n)} = \sum_{0\leq k\leq n} \left(\begin{array}{c} n \\ k \end{array}\right)\ f^{(k)}g^{(n-k)}$$

This formula is proved by induction w.r.t. the parameter n.

$$\begin{aligned}(fg)^{(n+1)} &= \sum_{0\leq k\leq n} \left(\begin{array}{c} n \\ k \end{array}\right)\ f^{(k+1)}g^{((n+1)-(k+1))} + \sum_{0\leq l\leq n} \left(\begin{array}{c} n \\ l \end{array}\right)\ f^{(l)}g^{(n+1-l)} \\ &= \sum_{1\leq l\leq n} \underbrace{\left[\left(\begin{array}{c} n \\ l-1 \end{array}\right) + \left(\begin{array}{c} n \\ l \end{array}\right)\right]}_{=\left(\begin{array}{c} n+1 \\ l \end{array}\right)}\ f^{(l)}g^{(n+1-l)} + \left(\begin{array}{c} n \\ n \end{array}\right) f^{(n+1)} + \left(\begin{array}{c} n \\ 0 \end{array}\right) g^{(n+1)}\end{aligned}$$

Applying this formula to $f(x) = -2x$ and $g(x) = e^{-x^2}$ we find that

$$\begin{aligned}\frac{d^{n+1}}{dx^{n+1}}e^{-x^2} &= \frac{d^n}{dx^n}(-2xe^{-x^2}) \\ &= \left(\begin{array}{c} n \\ 0 \end{array}\right)(-2x)\ \frac{d^n}{dx^n}e^{-x^2} + \left(\begin{array}{c} n \\ 1 \end{array}\right)(-2)\ \frac{d^{n-1}}{dx^{n-1}}e^{-x^2} \\ &= -2x\ \frac{d^n}{dx^n}e^{-x^2} - 2n\frac{d^{n-1}}{dx^{n-1}}e^{-x^2} \Rightarrow \mathbb{H}_{n+1}(x) = 2x\mathbb{H}_n(x) - 2n\mathbb{H}_{n-1}(x) \qquad (11.3)\end{aligned}$$

Combining (11.2) and (11.3), we have

$$2x\mathbb{H}_n(x) - 2n\mathbb{H}_{n-1}(x) = 2x\mathbb{H}_n(x) - \mathbb{H}'_n(x) \Rightarrow \mathbb{H}'_n = 2n\mathbb{H}_{n-1} \Rightarrow \mathbb{H}''_n = 2n\mathbb{H}'_{n-1}$$

and therefore

$$\begin{aligned}\mathbb{H}'_n(x) = 2x\mathbb{H}_n(x) - \mathbb{H}_{n+1}(x) \Rightarrow\ \mathbb{H}''_n(x) &= 2\mathbb{H}_n(x) + 2x\mathbb{H}'_n(x) - \mathbb{H}'_{n+1}(x) \qquad (11.4) \\ &= 2\mathbb{H}_n(x) + 2x\mathbb{H}'_n(x) - 2(n+1)\mathbb{H}_n(x) \\ &= 2x\mathbb{H}'_n(x) - 2n\mathbb{H}_n(x)\end{aligned}$$

We set

$$\overline{\mathbb{H}}_n(x) := e^{-x^2/2}\mathbb{H}_n(x) \quad \Longleftrightarrow \quad \mathbb{H}_n(x) = e^{x^2/2}\ \overline{\mathbb{H}}_n(x)$$

Using the fact that

$$\begin{aligned}\mathbb{H}'_n(x) &= e^{x^2/2}\ \left[x\overline{\mathbb{H}}_n(x) + \overline{\mathbb{H}}'_n(x)\right] \\ \mathbb{H}''_n(x) &= e^{x^2/2}\ \left(\left[\overline{\mathbb{H}}_n(x) + x\overline{\mathbb{H}}'_n(x) + \overline{\mathbb{H}}''_n(x)\right] + x\left[x\overline{\mathbb{H}}_n(x) + \overline{\mathbb{H}}'_n(x)\right]\right) \\ &= e^{x^2/2}\ \left[(1+x^2)\overline{\mathbb{H}}_n(x) + 2x\overline{\mathbb{H}}'_n(x) + \overline{\mathbb{H}}''_n(x)\right]\end{aligned}$$

we check that

$$\begin{aligned}(11.4) \Leftrightarrow\ 0 &= \mathbb{H}_n''(x) - 2x\mathbb{H}_n'(x) + 2n\mathbb{H}_n(x) \\ &= e^{x^2/2}\left[\left[(1+x^2)\overline{\mathbb{H}}_n(x) + 2x\overline{\mathbb{H}}_n'(x) + \overline{\mathbb{H}}_n''(x)\right] - 2x\left[x\overline{\mathbb{H}}_n(x) + \overline{\mathbb{H}}_n'(x)\right] + 2n\mathbb{H}_n(x)\right] \\ &= e^{x^2/2}\left[\overline{\mathbb{H}}_n''(x) + ((2n+1) - x^2)\overline{\mathbb{H}}_n(x)\right]\end{aligned}$$

and

$$\overline{\mathbb{H}}_n''(x) = (x^2 - (2n+1))\ \overline{\mathbb{H}}_n(x)$$

This shows that $\psi_n(x) = \overline{\mathbb{H}}_n(x)$ and therefore

$$\varphi_n(x) \propto \psi_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) = \overline{\mathbb{H}}_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) = e^{-\frac{x^2}{2}\frac{m\omega}{\hbar}}\ \mathbb{H}_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right)$$

Finally, we obtain the orthornormal basis of eigenfunctions by setting

$$\varphi_n(x) := \sqrt{\frac{1}{2^n n!\sqrt{\pi}}}\ \left(\frac{m\omega}{\hbar}\right)^{1/4} \exp\left[-\frac{x^2}{2}\frac{m\omega}{\hbar}\right]\ \mathbb{H}_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right)$$

11.1.3 Trial ground state energies

We notice that

$$L^V(\varphi_0) = L(\varphi_0) - V\varphi_0 = E_0\varphi_0 \Rightarrow V = \varphi_0^{-1}L(\varphi_0) - E_0$$

Using the exponential change of probability measures discussed in section 5.3.3, this implies that

$$\begin{aligned}\gamma_t(f) &:= \mathbb{E}\left(f(X_t)\ e^{-\int_0^t V(X_s)ds}\right) = e^{E_0 t}\ \mathbb{E}\left(f(X_t)\ e^{-\int_0^t [\varphi_0^{-1}L(\varphi_0)](X_s)ds}\right) \\ &= e^{E_0 t}\ \mathbb{E}\left(\frac{\varphi_0(X_0)}{\varphi_0(X_t)}\ f(X_t)\ \frac{\varphi_0(X_t)}{\varphi_0(X_0)}\ e^{-\int_0^t [\varphi_0^{-1}L(\varphi_0)](X_s)ds}\right) \\ &= e^{E_0 t}\ \eta_0(\varphi_0)\ \mathbb{E}\left(\varphi_0^{-1}(X_t^{\varphi_0})\ f(X_t^{\varphi_0})\right)\end{aligned}$$

and

$$\eta_t(f) := \frac{\gamma_t(f)}{\gamma_t(1)} = \frac{\mathbb{E}\left(\varphi_0^{-1}(X_t^{\varphi_0})\ f(X_t^{\varphi_0})\right)}{\mathbb{E}\left(\varphi_0^{-1}(X_t^{\varphi_0})\right)}$$

where $\eta_0 = \text{Law}(X_0)$, and $X_t^{\varphi_0}$ stands for the Markov process with initial distribution $\eta_0^{[\varphi_0]} = \Psi_{\varphi_0}(\eta_0)$ and infinitesimal generator

$$L^{[\varphi_0]}(f) = L(f) + \varphi_0^{-1}\Gamma_L(\varphi_0, f)$$

These stochastic models are the continuous time version of the discrete time Doob h-processes discussed in section 4.3.2. We also refer the reader to the end of section 5.3.3 for some work out examples of jump-diffusion processes with generator $L^{[\varphi_0]}$.

The ground state φ_0 is usually unknown and we often use the $\varphi_{\mathcal{T}}$-process $X_t^{\varphi_{\mathcal{T}}}$ associated with a trial energy function (a.k.a. guiding wave function) denoted by $\varphi_{\mathcal{T}}$. In this case, we have

$$\begin{aligned}\gamma_t(f) &:= \mathbb{E}\left(f(X_t)\ e^{-\int_0^t V(X_s)ds}\right) = \mathbb{E}\left(f(X_t)\ e^{-\int_0^t (V - [\varphi_{\mathcal{T}}^{-1}L(\varphi_{\mathcal{T}})])(X_s)ds} e^{-\int_0^t [\varphi_{\mathcal{T}}^{-1}L(\varphi_{\mathcal{T}})](X_s)ds}\right) \\ &= \eta_0(\varphi_{\mathcal{T}})\ \mathbb{E}\left(\varphi_{\mathcal{T}}^{-1}(X_t^{\varphi_{\mathcal{T}}})\ f(X_t^{\varphi_{\mathcal{T}}})\ e^{-\int_0^t (V - [\varphi_{\mathcal{T}}^{-1}L(\varphi_{\mathcal{T}})])(X_s^{\varphi_{\mathcal{T}}})ds}\right)\end{aligned}$$

The above formula is sometimes written as follows

$$\gamma_t(f) = \eta_0(\varphi_{\mathcal{T}}) \; \mathbb{E}\left(\varphi_{\mathcal{T}}^{-1}(X_t^{\varphi_{\mathcal{T}}}) \; f(X_t^{\varphi_{\mathcal{T}}}) \; e^{-\int_0^t V_{\mathcal{T}}(X_s^{\varphi_{\mathcal{T}}})ds}\right)$$

with the trial ground state energy (a.k.a. local energy) $V_{\mathcal{T}}$ given by

$$V_{\mathcal{T}}(x) := V(x) - \left[\varphi_{\mathcal{T}}^{-1} L(\varphi_{\mathcal{T}})\right](x) = \left[\varphi_{\mathcal{T}}^{-1} \mathcal{H}(\varphi_{\mathcal{T}})\right](x)$$

In the above display, $X_t^{\varphi_{\mathcal{T}}}$ stands for the $\varphi_{\mathcal{T}}$-twisted process with initial distribution $\eta_0^{[\varphi_{\mathcal{T}}]} = \Psi_{\varphi_{\mathcal{T}}}(\eta_0)$ and infinitesimal generator

$$L^{[\varphi_{\mathcal{T}}]}(f) = L(f) + \varphi_{\mathcal{T}}^{-1} \Gamma_L(\varphi_{\mathcal{T}}, f)$$

11.2 Signal Processing

11.2.1 Nonlinear filtering models

Suppose that at every time step the state of the Markov chain X_n taking values in some state spaces E_n^X is partially observed according to the following schematic picture

$$\begin{array}{ccccccc} X_0 & \longrightarrow & X_1 & \longrightarrow & X_2 & \longrightarrow & \dots \\ \downarrow & & \downarrow & & \downarrow & & \\ Y_0 & & Y_1 & & Y_2 & & \dots \end{array}$$

The typical model is given by a reference Markov chain model X_n and some partial and noisy observation Y_n. We denote by M_n the elementary Markov transitions of the Markov chain X_n. The pair process (X_n, Y_n) usually forms a Markov chain on some product measurable state space $\left(E_n^X \times E_n^Y\right)$ with elementary transitions given

$$\mathbb{P}\left((X_n, Y_n) \in d(x,y) \mid (X_{n-1}, Y_{n-1})\right) = M_n(X_{n-1}, dx) \; g_n(x,y) \; \lambda_n(dy) \tag{11.5}$$

for some positive likelihood function g_n, and some reference probability measure λ_n on E_n^Y. In the further development of this section, we fix the observation sequence $Y_n = y_n$, for $n \geq 0$. As traditional in nonlinear filtering literature, when there is no possible confusion, we slightly abuse the notation and we suppress as much as we can the dependence on the observation sequence. For any $n \geq 0$, we set

$$G_n(x_n) = g_n(x_n, y_n) \tag{11.6}$$

With a slight abuse of notation, sometimes we denote by $p(\boldsymbol{y_n}) := p(y_0, \dots, y_n)$ the density of the historical observation sequence

$$\boldsymbol{Y_n} := (Y_0, \dots, Y_n) \quad \text{w.r.t. the product measure} \quad \boldsymbol{\lambda_n(dy_n)} := \otimes_{0 \leq p \leq n} \lambda_p(dy_p)$$

that is, we have that

$$\mathbb{P}\left(\boldsymbol{Y_n} \in \boldsymbol{dy_n}\right) = p(\boldsymbol{y_n}) \; \boldsymbol{\lambda_n(dy_n)}$$

By construction, the conditional distribution of the historical process of the signal

$$\boldsymbol{X_n} := (X_0, \dots, X_n)$$

given the sequence of observations $\boldsymbol{Y_{n-1}} := (Y_0, \dots, Y_{n-1})$ coincides with the Feynman-Kac measures (4.17) associated with the pair (G_n, M_n); that is, we have that

$$\boldsymbol{\eta_n} = \text{Law}(\boldsymbol{X_n} \mid \boldsymbol{Y_{n-1}}) \tag{11.7}$$

In addition, the normalizing constants defined in (4.17) take the following form

$$\mathcal{Z}_{n+1} = p(\boldsymbol{y_n}) = \mathbb{E}\left(\prod_{0\leq k\leq n} G_k(x_k)\right)$$

To underline the dependence on the observation sequence, sometimes we write $\mathcal{Z}_{n+1}(\boldsymbol{y})$, the normalizing constants associated with a given sequence of observations $\boldsymbol{y} = (y_p)_{0\leq p\leq n}$.

In this context, the optimal one step predictor η_n and the optimal filter $\widehat{\eta}_n$ are given by the n-th time marginal distribution defined by

$$\eta_n \quad = \quad \mathrm{Law}\left(X_n \mid \boldsymbol{Y_{n-1}}\right) \qquad \widehat{\eta}_n = \Psi_{G_n}(\eta_n) = \mathrm{Law}\left(X_n \mid \boldsymbol{Y_{n-1}}\right) \tag{11.8}$$

11.2.2 Regulation problems

This section is concerned with the estimation of the conditional distributions of the signal-noise given the observations in terms of a genealogical tree based model. These mathematical objects are closely related to optimal regulation problems. In this situation, we have a dedicated controlled complex system and a given reference trajectory. The problem is to compute the optimal sequence of controls that minimizes some energy cost function and drives the system as close as possible to some reference trajectory.

These regulation problems can be interpreted in terms of the maximum likelihood of a dual filtering problem. Inversely, the logarithm of the conditional distributions of a filtering model in path space can be interpreted in terms of the cost associated with a given optimal regulation model. In this context the mean field IPS genealogical tree model can be interpreted as a genealogical tree based decision tree algorithm.

For a detailed discussion on these models, we refer the reader to the appendix of the book [390]. Several real-world application of these decision tree models to optimal thermal processing, w.r.t. specified temperature trajectories, and robotic optimal control problems, can be found in the series of articles [354, 355, 464].

We further assume the signal process given by recursive equations on some spaces E_n of the following form

$$X_n := F_n(X_{n-1}, U_n) \tag{11.9}$$

In the above display, U_n stands for a sequence of independent, and independent of X_0, random variables with distribution ν_n on some state spaces $\mathcal{U}_n$. We also assume that F_n is a measurable function from $(E_{n-1}\times\mathcal{U}_n)$ into E_n. For $n=0$, we set $X_0 = U_0 \in E_0 = \mathcal{U}_0$.

We denote by

$$X_n := \psi_n(U_0,\ldots,U_n)$$

the stochastic semigroup associated with the random system (11.9).

We let $\mathbb{Q}_n$ be the Feynman-Kac model (4.17) associated with the reference Markov chain $\mathcal{X}_n$ and the potential functions $\mathcal{G}_n$

$$\mathcal{X}_n := (U_0,\ldots,U_n) \quad \text{and} \quad \mathcal{G}_n := G_n\circ\psi_n$$

By construction, we have

$$\mathbb{Q}_n = \mathrm{Law}\left((U_0,(U_0,U_1),\ldots,(U_0,\ldots,U_n)) \mid \forall 0\leq p<n \quad Y_p = y_p\right)$$

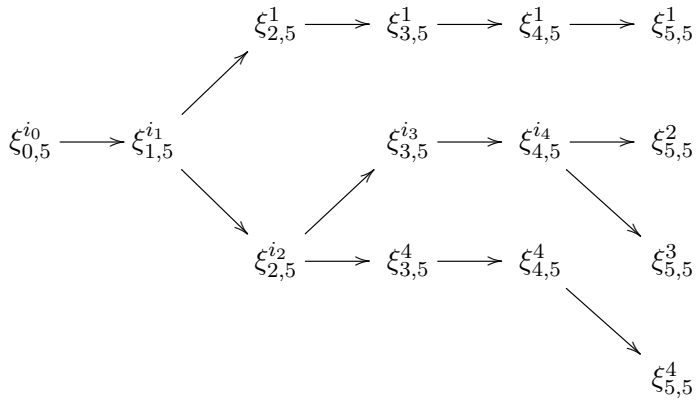
Thus, the n-th time marginal measures η_n are given by

$$\eta_n = \mathrm{Law}\left((U_0,\ldots,U_n) \mid \forall 0\leq p<n \quad Y_p = y_p\right)$$

These conditional distributions can be estimated using the genealogical tree based particle measures

$$\boldsymbol{\eta_n^N} := \frac{1}{N}\sum_{i=1}^{N}\delta_{\xi_n^i} \quad \text{with} \quad \xi_n^i := (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)$$

discussed in Section 10.2. The occupation measures of these ancestral trees are illustrated below for $(N,n) = (4,5)$, for any $i_0 \in \{1,2,3,4\}$, $i_2 \in \{2,3,4\}$, $i_3 \in \{2,3\}$, and any $i_4 \in \{2,3\}$:



The ancestral lines represent the conditional distribution of the sequence $(U_0, U_1, U_2, U_3, U_4, U_5)$ w.r.t. the sequence of observations $(Y_0, Y_1, Y_2, Y_3, Y_4)$, in terms of a likely initial condition $\xi_{0,5}^{i_0}$, and 4 likely signal-noise sequences $(\xi_{1,5}^i, \xi_{2,5}^i, \xi_{3,5}^i, \xi_{4,5}^i, \xi_{5,5}^i)_{i=1,2,3,4}$.

We further assume the random variables U_n have a density $h_n(u)$ w.r.t. to some reference distribution μ_n on $\mathcal{U}_n$, and we set $\boldsymbol{\mu_n} = \otimes_{0\leq p\leq n}\mu_p$. By construction, we have $\eta_n \ll \boldsymbol{\mu_n}$ and

$$\frac{d\eta_n}{d\boldsymbol{\mu_n}}(u_0,\ldots,u_n) \propto \left\{\prod_{0\leq p\leq n} h_p(u_p)\right\} \times \left\{\prod_{0\leq p<n} g_p(\psi_p(u_0,\ldots,u_p), y_p)\right\}$$

The log-likelihood function

$$V_n(u_0,\ldots,u_n) := -\log\left(\frac{d\eta_n}{d\boldsymbol{\mu_n}}(u_0,\ldots,u_n)\right)$$

can be rewritten as follows

$$V_n(u_0,\ldots,u_n) = \sum_{0\leq p\leq n} c_p(u_p) + \sum_{0\leq p<n} C_p\left(\psi_p(u_0,\ldots,u_p), y_p\right)$$

with the local logarithmic cost functions $(c_n, C_n) = (-\log h_n, -\log g_n)$. The function $-c_n$ can be interpreted as an energy type function on the control sequence u_n, and the energy function $-C_n$ measures the difference between the controlled semigroup of the system $\psi_n(u_0,\ldots,u_n)$ and the reference observation state y_n.

The optimal sequence of control that maximizes the log-likelihood function can be computed using the genealogical tree occupation measure. The estimate is defined by

$$\boldsymbol{\eta_n^N} - \text{ess inf } V_n := \min_{1\leq i\leq N} V_n(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \longrightarrow_{N\to\infty} \eta_n - \text{ess inf } V_n \quad \text{(in probability)}$$

11.2.3 Ensemble Kalman filters

We return to the linear Gaussian signal-observation model (8.24) discussed in section 8.3.1.

Given some probability measure η on $\mathbb{R}^d$, whenever they exist, we denote by m_η and P_η the mean value and the covariance matrix defined by

$$m_\eta = \int \varphi(x)\ \eta(dx) \quad \text{and} \quad P_\eta := \eta\left([\varphi - \eta(\varphi)]\ \ [\varphi - \eta(\varphi)]'\right)$$

with the column identity vector $\varphi(x) = x \in \mathbb{R}^p$. Using (8.27) and (8.28), and using the Gaussian-updating formula (8.29), the Kalman recursion (8.26) can alternatively be written as follows

$$\eta_n \xrightarrow{\text{e-updating}} \widetilde{\eta}_n := \widetilde{\Psi}(\eta_n) \xrightarrow{\text{prediction}} \eta_{n+1} = \widetilde{\eta}_n M_{n+1} \tag{11.10}$$

where $\widetilde{\Psi}(\eta_n) = \widetilde{\eta}_n$ stands for the distribution of the random variable

$$\widetilde{X}_n := \overline{X}_n + \mathbf{Gain}_{n,\eta_n}(y_n - C_n\overline{X}_n - c_n - D_nV_n)$$

with

$$\mathbf{Gain}_{n,\eta_n} = P_{\eta_n}C_n'(C_nP_{\eta_n}C_n' + D_nR_n^vD_n')^{-1} \quad \text{and} \quad \text{Law}(\overline{X}_n) = \eta_n\ .$$

In the above display formulae, V_n stands for a sequence of independent centered Gaussian random sequences with covariance matrices R_n^v. In other words, we have that

$$\widetilde{\Psi}(\eta_n)(f) = \int f\left[x + \mathbf{Gain}_{n,\eta_n}(y_n - C_nx - D_nv)\right]\ \eta_n(dx)\ \mathcal{N}(0, R_n^v)(dv)$$

For linear-Gaussian models, the evolution equations (11.10) and (8.26) are equivalent. Indeed, using the Gaussian-updating formula (8.29), if we consider the Gaussian likelihood functions

$$G_n(x) := g_{(C_nx+c_n, D_nR_n^vD_n')}(y_n)$$

discussed in (8.28), then we have

$$\eta_n = \mathcal{N}\left(m_{\eta_n}, P_{\eta_n}\right) \Longrightarrow \Psi_{G_n}\left(\eta_n\right) = \mathcal{N}(m_{\widetilde{\eta}_n}, P_{\widetilde{\eta}_n}) = \widetilde{\Psi}(\eta_n)$$

$$\text{with} \quad m_{\widetilde{\eta}_n} = m_{\eta_n} + \mathbf{Gain}_{n,\eta_n}\left(y_n - C_nm_{\eta_n} - c_n\right) \quad \text{and} \quad P_{\widetilde{\eta}_n} = (Id - \mathbf{Gain}_{n,\eta_n}C_n)P_{\eta_n}.$$

Clearly for non Gaussian models, the e-updating transition is not equivalent to the Bayes' rule. We consider the two step Markov chain model on $\mathbb{R}^p$ defined by the following synthetic diagram

$$\overline{X}_n \to \widetilde{X}_n = \overline{X}_n + \mathbf{Gain}_{n,\eta_n}(y_n - C_n\overline{X}_n - c_n - D_nV_n) \to \overline{X}_{n+1} = A_{n+1}\widetilde{X}_n + a_n + B_{n+1}W_{n+1}$$

with the initial condition $\overline{X}_0 = X_0$, and the same Gaussian r.v. (X_0, V_n, W_n) as the ones defined in (8.24). The elementary transition $\overline{X}_n \to \widetilde{X}_n$ is given by

$$\widetilde{S}_{n,\eta_n}(x_n, d\widetilde{x}_n) = \mathbb{P}\left(\widetilde{X}_n \in d\widetilde{x}_n \mid \overline{X}_n = x_n\right)$$

so that

$$\widetilde{\eta}_n = \widetilde{\Psi}(\eta_n) = \eta_n\widetilde{S}_{n,\eta_n}$$

This yields the McKean interpretation of the Kalman filter

$$\eta_{n+1} \quad = \quad \eta_nK_{n+1,\eta_n} \quad \text{with} \quad K_{n+1,\eta_n} := \widetilde{S}_{n,\eta_n}M_{n+1}$$

These discrete generation McKean models are discussed in Section 9.1. The mean field particle model (9.28) associated with this McKean model is defined by evolving an $(\mathbb{R}^p)^N$-valued and two step Markov chain model

$$\xi_n := \left(\xi_n^i\right)_{1\leq i\leq N} \xrightarrow{\widetilde{S}_{n,\eta_n^N}} \widetilde{\xi}_n := \left(\widetilde{\xi}_n^i\right)_{1\leq i\leq N} \xrightarrow{M_{n+1}} \xi_{n+1} := \left(\xi_{n+1}^i\right)_{1\leq i\leq N}$$

with $\eta_n^N := \frac{1}{N}\sum_{1\leq i\leq N}\delta_{\xi_n^i}$. More formally, the elementary transitions (9.28) are decomposed into the following steps

$$\xi_n^i \to \widetilde{\xi}_n^i = \xi_n^i + \mathbf{Gain}_{n,\eta_n^N}(y_n - C_n\xi_n^i - c_n - D_nV_n^i) \to \xi_{n+1}^i = A_{n+1}\widetilde{\xi}_n^i + a_n + B_{n+1}W_{n+1}^i$$

for any $i \in \{1,\ldots,N\}$, with i.i.d. copies $(V_n^i, W_n^i)_{1\leq i\leq N}$ of (V_n, W_n), and the empirical approximation of the covariance function

$$P_{\eta_n^N} := \eta_n^N\left(\left[\varphi - \eta_n^N(\varphi)\right]\quad\left[\varphi - \eta_n^N(\varphi)\right]'\right)$$

The main advantage of this mean field formulation of the Kalman filter comes from the fact that for large dimension signals intractable covariance prediction error matrices P_n^- are now computed using the sampled mean field particle empirical matrices $P_{\eta_n^N}$.

Mimicking (8.30), the density $p_n(y_0,\ldots,y_n)$ of the sequence of observation $(Y_0,\ldots,Y_n)$ evaluated at the random observation path $(Y_0,\ldots,Y_n)$ can be approximated by

$$p_n^N(Y_0,\ldots,Y_n) = \prod_{k=0}^{n} g_{\left(C_k\widehat{X}_k^{N,-}+c_k,\ \Sigma_k\left(P_{\eta_k^N}\right)\right)}(Y_k) \quad \text{with} \quad \widehat{X}_k^{N,-} = \frac{1}{N}\sum_{1\leq i\leq N}\xi_k^i \tag{11.11}$$

These mean field approximations are widely used in meteorological forecasting problems, where the signal process comes from a grid type approximation of Navier-Stokes' partial different equations arising in fluid mechanics.

Further details of these mean field particle filters, and their performance analysis, can be found in the recent articles by F. Le Gland, V. Monbet, and V. D. Tran [405, 406], as well as in the Ph.D. thesis of V. D. Tran [573] in 2009, and the one by Ch. Baehr [29] in 2008. We also refer the reader to the pioneering work by G. Evensen on ensemble Kalman filters [259, 260, 261], and a series of articles [11, 91, 341] on the numerical performance of these models in forecasting data assimilation problems.

11.2.4 Approximate Bayesian computation

As their name indicates, Approximate Bayesian Computation (*abbreviated ABC*) are Bayesian inference methods used to evaluate posterior distributions without having to calculate likelihoods. For instance, in biology applications and more particularly in predictive bacteriology and food risk analysis, the observations of a kinetic biological complex system are given by counting bacteria individuals after successive dilutions of a food sample coming from an in vitro culture [256, 257, 281, 282]. Of course, this experimental observation process is often modeled by a series of Poisson type dependent random variables but the computation of the likelihood function often requires successive summations over the set of all the integers. In this situation likelihood functions are computationally intractable, or too costly to estimate in a reasonable time.

One of the central ideas of ABC methods is to replace the evaluation of the likelihood function by a simulation-based procedure of the observation process coupled with a numerical comparison between the observed and simulated data. This strategy is rather well known in particle filtering literature; see for instance [163, 164, 165]. In the same vein, these additional levels of simulation-based approximations can also be extended to compute the posterior distribution of fixed parameters in hidden Markov chain models. In signal processing literature, these ABC type mean field IPS models are sometimes called convolution particle filters; see for instance [95, 96, 515, 580].

More formally, in some instance the likelihood functions $x_n \mapsto g_n(x_n, y_n)$ in (11.6) are computationally intractable, or too expensive to evaluate in a reasonable computational time. To solve this problem, a natural solution is to sample pseudo-observations. The central idea is to sample the signal-observation Markov chain $\overline{X}_n = (X_n, Y_n) \in E_n^{\overline{X}} = (E_n^X \times E_n^Y)$, and compare the values of the sampled observations with the real observations sequence.

To describe with some conciseness these approximate filtering models, we notice that the transitions of $\overline{X}_n$ are given by

$$\overline{M}_n(\overline{X}_{n-1}, d(x_n, y_n)) = M_n(X_{n-1}, dx_n)\ g_n(x_n, y_n)\ \lambda_n(dy_n)$$

To simplify the presentation, we further assume that $E_n^Y = \mathbb{R}^d$, for some $d \geq 1$, and we let g be a Borel bounded nonnegative function on $\mathbb{R}^d$ such that

$$\int g(u)\ du = 1 \quad \int u\ g(u)\ du = 0 \quad \text{and} \quad \int |u|^3\ g(u)\ du < \infty$$

Then, we set for any $\epsilon > 0$, any $\overline{x} = (x, y) \in E_n^{\overline{X}}$, and $z \in \mathbb{R}^d$

$$g_{\epsilon,n}(\overline{x}, z) = g_{\epsilon,n}((x, y), z) = \epsilon^{-d}\ g\left((y - z)/\epsilon\right)$$

Finally, we consider a Markov chain $(\overline{X}_n, Y_n^\epsilon)$ on the augmented state space $\left(E_n^{\overline{X}} \times E_n^Y\right)$ with transitions given

$$\mathbb{P}\left((\overline{X}_n, Y_n^\epsilon) \in d(\overline{x}_n, y_n) \mid (\overline{X}_{n-1}, Y_{n-1}^\epsilon)\right) = \overline{M}_n(\overline{X}_{n-1}, d\overline{x}_n)\ g_{\epsilon,n}(\overline{x}_n, y_n)\ dy_n \tag{11.12}$$

This approximated filtering problem has exactly the same form as the one introduced in (11.5). In this situation, the mean field particle model is defined in terms of signal-observation valued particles, and the selection potential function is given by the pseudo-likelihood functions $g_{\epsilon,n}(., y_n)$, where y_n stands for the value of the observation sequence at time n.

11.2.5 Quenched and annealed filters

Suppose that at every time step the state of a Markov chain with two coordinates (Θ_n, X_n) is partially observed according to the following schematic picture

$$\begin{array}{ccccccc} \Theta_0 & \longrightarrow & \Theta_1 & \longrightarrow & \Theta_2 & \longrightarrow & \ldots \\ \downarrow & & \downarrow & & \downarrow & & \\ X_0 & \longrightarrow & X_1 & \longrightarrow & X_2 & \longrightarrow & \ldots \\ \downarrow & & \downarrow & & \downarrow & & \\ Y_0 & & Y_1 & & Y_2 & & \ldots \end{array} \tag{11.13}$$

We assume that the Markov chain model (Θ_n, X_n) evolves on some state spaces $(\Xi_n \times E_n)$ with Markov transitions of the form (10.31). Given a realization of the Markov chain $\Theta_n = \theta_n$, and the random state $X_n = x_n$, the observation Y_n is a random variable taking values on some finite state space E_n^Y, with distribution

$$\mathbb{P}\left(Y_n \in dy_n \mid X_n = x_n, \Theta_n = \theta_n\right) := g_{\theta_n,n}(x_n, y_n)\ \lambda_n(dy_n) \tag{11.14}$$

for some reference positive measure λ_n on E_n^Y. We fix a sequence of observations $\boldsymbol{y} = (y_n)_{n\geq 0}$, and for any realization $\boldsymbol{\theta} = (\theta_n)_{n\geq 0}$ of the chain $\boldsymbol{\Theta} = (\Theta_n)_{n\geq 0}$ we set

$$G_{\theta_n,n}(x_n) = g_{\theta_n,n}(x_n, y_n) \tag{11.15}$$

This model coincides with the quenched and annealed Feynman-Kac models (10.34) discussed in section 10.6.2.

More precisely, the normalized version $\mathbb{Q}_{\boldsymbol{n}}$ of the measure $\boldsymbol{\Gamma_n}$ defined in (10.35) coincides with the conditional distribution of the historical process of the signal $\boldsymbol{X_n} := (X_0, \ldots, X_n)$ given the sequence of observations $\boldsymbol{Y_{n-1}}$. More formally, we have that

$$\boldsymbol{\eta_n} = \mathbb{Q}_n = \mathrm{Law}(\boldsymbol{X_n} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}})$$

In addition, for any realization of the historical process

$$\boldsymbol{\Theta_n} = (\Theta_p)_{0\leq p\leq n} = (\theta_p)_{0\leq p\leq n} = \boldsymbol{\theta_n} \in \boldsymbol{\Xi_n} = \prod_{0\leq p\leq n} \Xi_p$$

the quenched Feynman-Kac distributions (10.33) coincide with the quenched conditional distributions

$$\begin{aligned} \mathbb{Q}_{\boldsymbol{\theta_n},n} &= \mathrm{Law}(\boldsymbol{X_n} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}},\ \boldsymbol{\Theta_n} = \boldsymbol{\theta_n}) \\ \eta_{\boldsymbol{\theta_n},n} &= \mathrm{Law}(X_n \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}},\ \boldsymbol{\Theta_n} = \boldsymbol{\theta_n}) \\ \widehat{\eta}_{\boldsymbol{\theta_n},n} &= \Psi_{G_{\theta_n,n}}(\eta_{\boldsymbol{\theta_n},n}) = \mathrm{Law}(X_n \mid \boldsymbol{Y_n} = \boldsymbol{y_n},\ \boldsymbol{\Theta_n} = \boldsymbol{\theta_n}) \end{aligned} \tag{11.16}$$

In this situation, the normalizing constants of the Feynman-Kac measures $\mathbb{Q}_{\boldsymbol{\theta_n},n}$ are given by

$$p(\boldsymbol{y_n} \mid \boldsymbol{\theta_n}) = \prod_{0\leq k\leq n} \boldsymbol{h_k(\theta_k)} \quad \text{with} \quad \boldsymbol{h_k(\theta_k)} := \eta_{\boldsymbol{\theta_k},k}\,(G_{\theta_k,k}) \tag{11.17}$$

where $p(\boldsymbol{y_n} \mid \boldsymbol{\theta_n})$ stands for the conditional density of the historical observation sequence $\boldsymbol{Y_n}$ w.r.t. the product measure $\boldsymbol{\lambda_n}$, given a realization of the historical process $\boldsymbol{\Theta_n} = \boldsymbol{\theta_n}$. Finally, using Bayes' rule we show that the posterior distribution of $\boldsymbol{\Theta_n}$ given $\boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}$ coincides with the annealed measure $\boldsymbol{\mu_n}$ defined in (10.36); that is, we have that

$$\boldsymbol{\mu_n} = \mathrm{Law}(\boldsymbol{\Theta_n} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}) \tag{11.18}$$

By (10.40), we also have that

$$\boldsymbol{F_n(\theta_n)} := \eta_{\boldsymbol{\theta_n},n}(f_n) \Rightarrow \boldsymbol{\mu_n(F_n)} = \mathbb{E}(f_n(X_n) \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}})$$

Finally, when the Markov transitions $M_{\theta_n,n}$ are absolutely continuous with respect to some measures λ_n on E_n, we have

$$\boldsymbol{F^{\#}_{f_n}(\theta_n)} := \mathbb{Q}_{\boldsymbol{\theta_n},n}(\boldsymbol{f_n}) = \boldsymbol{\eta_{\theta_n,n}(f_n)} \Rightarrow \boldsymbol{\mu_n(F_n)} = \mathbb{E}(\boldsymbol{f(X_n)} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}})$$

In the above displayed formula, $\mathbb{Q}_{\boldsymbol{\theta_n},n}$ stands for the backward Markov chain measure (10.41)

We illustrate the abstract models presented in (11.13) with a class of mean field type interacting Kalman filter. In signal processing literature, and Bayesian inference, these particle approximations are often referred to as particle methods in path space, or as Rao-Blackwellized particle filters (see for instance [172, 186, 234], and references therein).

We consider a Markov chain Θ_n taking values in some measurable state space Ξ_n, and a collection of matrices

$$A_n(\theta), B_n(\theta), C_n(\theta), D_n(\theta) \tag{11.19}$$

and vectors $a_n(\theta), b_n(\theta)$, indexed by $\theta \in \Xi_n$, of the same dimension as the matrices (A_n, B_n, C_n, D_n) and the vectors (a_n, b_n) introduced in (8.24). We let (Θ_n, X_n, Y_n) be the $(E \times \mathbb{R}^{p+q})$-valued Markov chain defined by the same recursive relations as in (8.24) by replacing (A_n, B_n, C_n, D_n) and (a_n, c_n) by $(A_n(\Theta_n), B_n(\Theta_n), C_n(\Theta_n), D_n(\Theta_n))$ and $(a_n(\Theta_n), c_n(\Theta_n))$.

We let $g_{(m,R)}(x)$, be the Gaussian densities associated with a mean and covariance matrix (m, R) introduced in (8.11). In this notation, the likelihood functions given in (11.15) and (11.14) are given by

$$G_{n,\theta_n}(x) = g_{(m_{\theta_n,n}(x),R_{\theta_n,n})}(y_n)$$

with

$$m_{\theta_n,n}(x) := C_n(\theta_n)\ x + c_n(\theta_n) \quad \text{and} \quad R_{\theta_n,n} := D_n(\theta_n) R_n^v D_n'(\theta_n)$$

In addition, given a realization of the historical process $\boldsymbol{\Theta} = (\Theta_n)_{n\geq 0}$ we have

$$\eta_{\boldsymbol{\theta_n},n} = \mathcal{N}(\widehat{X}_n^{\boldsymbol{\theta_n},-}, P_n^{\boldsymbol{\theta_n},-}) \quad \text{and} \quad \Psi_{G_{\theta_n,n}}(\eta_{\boldsymbol{\theta_n},n}) = \mathcal{N}(\widehat{X}_n^{\boldsymbol{\theta_n}}, P_n^{\boldsymbol{\theta_n}})$$

with some parameters $(\widehat{X}_n^{\boldsymbol{\theta_n},-}, P_n^{\boldsymbol{\theta_n},-})$ and $(\widehat{X}_n^{\boldsymbol{\theta_n}}, P_n^{\boldsymbol{\theta_n}})$ that can be computed using the same Kalman recursions (8.26) by replacing $(A_n, a_n, B_n, C_n, c_n, D_n)$ by $(A_n(\theta_n), a_n(\theta_n), B_n(\theta_n), C_n(\theta_n), c_n(\theta_n), D_n(\theta_n))$.

Using (8.30), we also easily check the following multiplicative formula

$$p(\boldsymbol{y_n}|\ \boldsymbol{\theta_n}) \quad = \quad \prod_{k=0}^{n} g_{\left(m_{\theta_k,k}\left(\widehat{X}_k^{\boldsymbol{\theta_k},-}\right),\Sigma_{\theta_k,k}\left(P_k^{\boldsymbol{\theta_k},-}\right)\right)}(y_k) \tag{11.20}$$

with

$$\Sigma_{\theta_k,k}(P_k^{\boldsymbol{\theta_k},-}) := C_k(\theta_k)P_k^{\boldsymbol{\theta_k},-}C_k'(\theta_k) + D_k(\theta_k)R_k^v D_k'(\theta_k)$$

In a more synthetic form, we have

$$p(\boldsymbol{y_n}|\ \boldsymbol{\theta_n}) = \prod_{k=0}^{n} \boldsymbol{h_k(\theta_k)}$$

with the functions

$$\boldsymbol{h_k(\theta_k)} := g_{(m_{\theta_k,k}(\widehat{X}_k^{\boldsymbol{\theta_k},-}),\Sigma_{\theta_k,k}(P_k^{\boldsymbol{\theta_k},-}))}(y_k) \in \mathbb{R}_+ \tag{11.21}$$

Using (11.18), the posterior distribution of $\boldsymbol{\Theta_n}$ given $\boldsymbol{Y_{n-1} = y_{n-1}}$ is given for any $F_n \in \mathcal{B}(\boldsymbol{\Xi_n})$ by the Feynman-Kac formula

$$\boldsymbol{\mu_n(F_n)} \quad = \quad \mathbb{E}(\boldsymbol{F_n(\Theta_n)} \mid \boldsymbol{Y_{n-1} = y_{n-1}}) \propto \mathbb{E}\left(\boldsymbol{F_n(\Theta_n)} \prod_{0\le p<n} \boldsymbol{h_p(\Theta_p)}\right)$$

By (10.40), we also have that

$$\boldsymbol{F_n(\theta_n)} := \eta_{\boldsymbol{\theta_n},n}(f_n) \Rightarrow \boldsymbol{\mu_n(F_n)} = \mathbb{E}(f_n(X_n) \mid \boldsymbol{Y_{n-1} = y_{n-1}})$$

The mean field particle approximation of these Feynman-Kac measures can be interpreted as a sequence of interacting Kalman filters. The model obtained by replacing the Kalman integration by particle filter integration coincides with the particle quenched and annealed models discussed in section 10.6.3.

11.3 Bayesian statistical inference

11.3.1 Hidden Markov chain models

We return to the quenched and annealed filtering models (11.13) discussed in section 11.2.5. We further assume that $\Theta_n = \Theta$, for any $n \ge 0$, where Θ stands for a r.v. with distribution λ on some measurable state space Ξ. In this situation, the measures $\mathbb{Q}_{\boldsymbol{\theta_n},n}$, and $\eta_{\boldsymbol{\theta_n},n}$ defined in (11.16) reduce to

$$\mathbb{Q}_{\theta,n} = \mathrm{Law}(\boldsymbol{X_n} \mid \boldsymbol{Y_{n-1} = y_{n-1}},\ \Theta = \theta) \quad \text{and} \quad \eta_{\theta,n} = \mathrm{Law}(X_n \mid \boldsymbol{Y_{n-1} = y_{n-1}},\ \Theta = \theta) \tag{11.22}$$

In addition, the normalizing constants of the Feynman-Kac measures $\mathbb{Q}_{\theta,n}$ are now given by

$$p(\boldsymbol{y_n} \mid \theta) = \prod_{0\le k\le n} h_k(\theta) \quad \text{with} \quad h_k(\theta) := \eta_{\theta,k}\left(G_{\theta,k}\right)$$

so that

$$\mu_n(d\theta) := \mathbb{P}(\Theta \in d\theta \mid \boldsymbol{Y_{n-1} = y_{n-1}}) \propto \left\{\prod_{0\le k\le n} h_k(\theta)\right\} \lambda(d\theta) := \nu_n(d\theta) \tag{11.23}$$

For the linear Gaussian models discussed in (11.19), the functions $h_k(\theta)$ coincides with the functions (11.21) and they are computed using the Kalman recursions associated with a given parameter θ.

11.3.2 Extended posterior distributions

For nonlinear models, we use the particle quenched and annealed models developed in section 10.6.3. To be more precise, we let

$$\overline{\Theta}_n = (\Theta, \chi_n) \in \overline{\Xi}_n := \left(\Xi \times \prod_{0 \le p \le n} E_p^N \right) \quad \text{with} \quad \chi_n = (\xi_0, \ldots, \xi_n)$$

where ξ_n stands for the conditional N-mean field particle model given the parameter Θ. We denote by $\mathcal{P}^{(n)}(\theta, dx_n)$ the conditional distribution of the N-particle model χ_n given a realization $\Theta = \theta$ of the parameter; that is

$$\mathcal{P}^{(n)}(\theta, dx_n) := \mathbb{P}\left(\chi_n \in dx_n \mid \Theta = \theta\right)$$

We also set

$$\overline{h}_n(\overline{\Theta}_n) = \eta_{\Theta,n}^N(G_{\theta,n})$$

In this notation, the formula (10.44) reduces to

$$\mathbb{E}\left(F(\Theta) \prod_{0 \le k < n} G_{\Theta,k}(X_k) \right) = \mathbb{E}\left(F(\Theta) \prod_{0 \le k < n} \overline{h}_k(\overline{\Theta}_k) \right) = \overline{\nu}_n(F \otimes 1)$$

for any $F \in \mathcal{B}(\Xi)$, with the measures $\overline{\nu}_n$ on $\overline{\Xi}_n$ defined for any $\overline{F}_n \in \mathcal{B}(\overline{\Xi}_n)$ by

$$\overline{\nu}_n(\overline{F}_n) = \mathbb{E}\left(\overline{F}_n(\overline{\Theta}_n) \prod_{0 \le k < n} \overline{h}_k(\overline{\Theta}_k) \right) \quad \text{and} \quad \overline{\mu}_n(\overline{F}_n) = \overline{\nu}_n(\overline{F}_n)/\overline{\nu}_n(1) \tag{11.24}$$

Using Bayes' rule we conclude that

$$\overline{\nu}_n(F \otimes 1) = \mathbb{E}\left(F(\Theta) \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}\right)$$

In other words, the Θ-marginal of $\overline{\nu}_n$ coincides with the posterior distribution of Θ given the sequence of observations $\boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}$ (also given by (11.23)).

In much the same way, equation (10.47) takes the form

$$\begin{aligned} \mathbb{E}\left(f_n(X_n) \prod_{0 \le k < n} G_{\Theta,k}(X_k) \right) &= \mathbb{E}\left(F_{f_n}(\Theta) \prod_{0 \le k < n} h_k(\Theta) \right) \\ &= \mathbb{E}\left(\overline{F}_{f_n}(\overline{\Theta}_n) \prod_{0 \le p < n} \overline{h}_p(\overline{\Theta}_p) \right) = \overline{\nu}_n(\overline{F}_{f_n}) \end{aligned} \tag{11.25}$$

with the functions

$$F_{f_n}(\theta) := \eta_{\theta,n}(f_n) \quad \text{and} \quad \overline{F}_{f_n}(\overline{\Theta}_n) := \eta_{\Theta,n}^N(f_n)$$

11.3.3 Particle Metropolis-Hasting model

We fix a time horizon n and we set

$$\forall 0 \le k \le n \quad \overline{h}_k^{(n)}(\Theta, \chi_n) = \eta_{\Theta,k}^N(G_{\theta,k})$$

In this notation, for any $0 \le k \le n$ we have

$$\overline{F}_k^{(n)}(\Theta, \chi_n) = \overline{F}_k(\Theta, \chi_k) \quad \Rightarrow \quad \overline{\nu}_k(\overline{F}_k) = \mathbb{E}\left(\overline{F}_k^{(n)}(\Theta, \chi_n) \prod_{0 \le l < k} \overline{h}_k^{(n)}(\Theta, \chi_n) \right) = \overline{\nu}_k^{(n)}(\overline{F}_k^{(n)})$$

with the product measures $\overline{\nu}_k^{(n)}$ defined by

$$\overline{\nu}_k^{(n)}(d\overline{\theta}) = \left\{ \prod_{0\leq l<k} \overline{h}_k^{(n)}(\overline{\theta}) \right\} \overline{\lambda}^{(n)}(d\overline{\theta}) \quad \text{with the reference measure} \quad \overline{\lambda}^{(n)} = \text{Law}\,(\Theta, \chi_n)$$

We consider a Markov transition $P(\theta, d\theta')$ on the parameter space s.t.

$$\lambda(d\theta)P(\theta, d\theta') \sim \lambda(d\theta')P(\theta', d\theta)$$

We associate with this exploration model the collection Markov chain transitions $K_k^{(n)}$, $1 \leq k \leq n$, on $\overline{\Xi}_n$ defined by

$$\forall \overline{\theta} = (\theta, \mathrm{x}_n)\,,\ \overline{\theta}' = \left(\theta', \mathrm{x}'_n\right) \in\in \overline{\Xi}_n \qquad K_k^{(n)}(\overline{\theta}, d\overline{\theta}') := P(\theta, d\theta')\ \mathcal{P}^{(n)}(\theta', d\mathrm{x}'_n)$$

By construction, we have

$$\overline{\nu}_k^{(n)}(d\overline{\theta}')\ K_k^{(n)}(\overline{\theta}', d\overline{\theta}) \ \sim\ \overline{\nu}_k^{(n)}(d\overline{\theta})\ K_k^{(n)}(\overline{\theta}, d\overline{\theta}')$$

with the Radon Nikodym derivatives

$$\begin{aligned} \frac{\overline{\nu}_k^{(n)}(d\overline{\theta}')\ K_k^{(n)}(\overline{\theta}', d\overline{\theta})}{\overline{\nu}_k^{(n)}(d\overline{\theta})\ K_k^{(n)}(\overline{\theta}, d\overline{\theta}')} &= \frac{\left\{\prod_{0\leq l<k} \overline{h}_k^{(n)}(\overline{\theta}')\right\}\ \overline{\lambda}^{(n)}(d\overline{\theta}')\ P(\theta', d\theta)\ \mathcal{P}^{(n)}(\theta, d\mathrm{x}_n)}{\left\{\prod_{0\leq l<k} \overline{h}_k^{(n)}(\overline{\theta})\right\}\ \overline{\lambda}^{(n)}(d\overline{\theta})\ P(\theta, d\theta')\ \mathcal{P}^{(n)}(\theta', d\mathrm{x}'_n)} \\ &= \left\{ \prod_{0\leq l<k} \frac{\overline{h}_k^{(n)}(\overline{\theta}')}{\overline{h}_k^{(n)}(\overline{\theta})} \right\} \times \frac{\lambda(d\theta')\ P(\theta', d\theta)}{\lambda(d\theta)\ P(\theta, d\theta')} \end{aligned}$$

The last assertion comes from the fact that

$$\overline{\lambda}^{(n)}(d\overline{\theta}) = \lambda(d\theta)\ \mathcal{P}^{(n)}(\theta, d\mathrm{x}_n)$$

$$\Rightarrow \overline{\lambda}^{(n)}(d\overline{\theta})\ P(\theta, d\theta')\ \mathcal{P}^{(n)}(\theta', d\mathrm{x}'_n) = \lambda(d\theta)\ P(\theta, d\theta')\ \left[\mathcal{P}^{(n)}(\theta, d\mathrm{x}_n)\ \mathcal{P}^{(n)}(\theta', d\mathrm{x}'_n)\right]$$

11.3.4 Particle Gibbs samplers

We return to the HMM model discussed in section 11.3.1. We fix the time horizon n. In view of (11.22) and (11.23) we have

$$\pi(d(\theta, \boldsymbol{x})) \ := \ \mathbb{P}\left((\Theta, \boldsymbol{X_n}) \in d(\theta, \boldsymbol{x}) \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}\right) = \mu_n(d\theta) \times \mathbb{Q}_{\theta,n}(\boldsymbol{dx_n})$$

This disintegration property can be rewritten as follows

$$\pi(d(\theta, \boldsymbol{x})) = \pi_1(d\theta)\ \pi_{1,2}(\theta, \boldsymbol{dx})$$

with

$$\begin{aligned} \pi_1(d\theta) \ &:= \ \mu_n(d\theta) = \mathbb{P}\left(\Theta \in d\theta \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}\right) \\ \pi_{1,2}(\theta, \boldsymbol{dx}) \ &:= \ \mathbb{Q}_{\theta,n}(\boldsymbol{dx}) = \mathbb{P}\left(\boldsymbol{X_n} \in d\boldsymbol{x} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}},\ \Theta = \theta\right) \end{aligned}$$

We further assume that

$$\forall (\theta, \theta') \in \Xi^2 \qquad \mathbb{Q}_{\theta,n} \ \sim\ \mathbb{Q}_{\theta',n}$$

This rather weak condition is satisfied as soon as $\eta_{\theta,0} \sim \eta_{\theta',0}$ and $M_{\theta,n}(x_{n-1}, .) \sim M_{\theta',n}(x_{n-1}, .)$, for any $(\theta, \theta') \in \Xi^2$ and any $x_{n-1} \in E_{n-1}$.

In this situation, we have the dual disintegration property

$$\pi(d(\theta, \boldsymbol{x})) = \pi_2(\boldsymbol{dx})\ \pi_{2,1}(\boldsymbol{x}, d\theta)$$

with

$$\begin{aligned} \pi_2 &:= \mu_n \mathbb{Q}_{\theta,n} = \mathrm{Law}(\boldsymbol{X_n} \mid \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}) \\ \pi_{2,1}(\boldsymbol{x}, d\theta) &:= \mu_n(d\theta)\ \frac{d\mathbb{Q}_{\theta,n}}{d\mu_n \mathbb{Q}_{\theta,n}}(\boldsymbol{x}) = \mathbb{P}\left(\Theta \in d\theta \mid \boldsymbol{X_n} = \boldsymbol{x}, \boldsymbol{Y_{n-1}} = \boldsymbol{y_{n-1}}\right) \end{aligned}$$

The design of MCMC samplers of these models have been discussed in section 7.6. For instance, let us choose a couple of Markov transitions

$$K_{\theta,1}(\boldsymbol{x}, \boldsymbol{dx'}) \quad \text{and} \quad K_{\boldsymbol{x'},2}(\theta, d\theta')$$

such that

$$\pi_{1,2}(\theta, \boldsymbol{dx}) = \int \pi_{1,2}(\theta, \boldsymbol{dx'})\ K_{\theta,1}(\boldsymbol{x'}, \boldsymbol{dx}) \tag{11.26}$$

$$\pi_{2,1}(\boldsymbol{x}, d\theta) = \int \pi_{2,1}(\boldsymbol{x}, d\theta')\ K_{\boldsymbol{x},2}(\theta', d\theta) \tag{11.27}$$

In this situation, π is an invariant distribution of the Markov transition

$$K((\theta, \boldsymbol{x}), d(\theta', \boldsymbol{x'}))) := K_{\theta,1}(\boldsymbol{x}, \boldsymbol{dx'})\ K_{\boldsymbol{x'},2}(\theta, d\theta')$$

For filtering problems with conjugate priors, the target measure $\pi_{2,1}(\boldsymbol{x}, d\theta)$ can be sampled directly using the conditional formulae developed in section 8.1.2. On the other hand, we can use the particle MCMC transition $K_{\theta,1}(\boldsymbol{x}, \boldsymbol{dx'})$ with frozen trajectory $\boldsymbol{x}$ developed in section 10.5.2.

11.3.5 Expected maximization models

We consider some parameter $\theta \in \mathbb{R}^d$ a Markov chain X_n, with elementary transitions $M_{n,\theta}$ on some measurable state spaces E_n with initial distribution $\eta_{0,\theta}$. We also consider a sequence of positive and bounded potential functions $G_{n,\theta}$ on the set E_n. We denote by $\mathbb{Q}_{n,\theta}$ the Feynman-Kac path measures (4.17) associated with the pairs $(M_{n,\theta}, G_{n,\theta})$.

We further assume that $\eta_{0,\theta} \ll \lambda_0$, and $M_{n,\theta}(x, .) \ll \lambda_n$, for some $\lambda_n \in \mathcal{M}_+(E_n)$ and we have

$$H_{0,\theta} = G_0 \times d\eta_{0,\theta}/d\lambda_0 > 0 \quad \text{and} \quad H_{n,\theta}(x, .) := G_{n-1}(x) \times dM_{n,\theta}(x, .)/d\lambda_n > 0$$

By construction, we have $\mathbb{Q}_{n,\theta} \ll \boldsymbol{\lambda_n} := \otimes_{0 \leq p \leq n} \lambda_p$ and the Radon-Nikodym derivates are given by the multiplicative formula

$$\frac{d\mathbb{Q}_{n,\theta}}{d\boldsymbol{\lambda_n}}(x_0, \ldots, x_n) := \mathbb{H}_{n,\theta}(x_0, \ldots, x_n) = \frac{1}{\mathcal{Z}_{n,\theta}} \prod_{0 \leq p \leq n} H_{\theta,p}(x_{p-1}, x_p)$$

for some normalizing constant $\mathcal{Z}_{n,\theta}$, with the convention $\boldsymbol{H}_{0,\theta}(x_{-1}, x_0) = H_{0,\theta}(x_0)$, for $n = 0$.

These models arise in various scientific disciplines. The prototype of model we have in mind is the parameter estimation in Hidden Markov chain problem discussed in Section 11.3.1. In this situation, we are given a pair of signal-observation processes that depend on some random parameter Θ. The distributions $\mathbb{Q}_{n,\theta}$ represent the conditional distribution of the random states, given a realization of the parameter $\Theta = \theta$, and their normalizing constants $\mathcal{Z}_{n,\theta}$ coincide with the distribution of the observations given $\Theta = \theta$. In this context, we are given a series of observation data related to some unknown θ and we want to maximize the mapping $\theta \mapsto \mathcal{Z}_{n,\theta}$ so that to find the parameter θ that "explains" these data.

One way to solve this problem is to use the celebrated expected maximization model. This statistical search model is a recursive gradient type algorithm that improves sequentially its solution computing the parameter that maximizes the expected log-likelihood function.

We briefly recall the principle of this gradient based approach. For any pair of parameters (θ, θ') we find that

$$\mathrm{Ent}\left(\mathbb{Q}_{n,\theta} | \, \mathbb{Q}_{n,\theta'}\right) \geq 0 \Rightarrow \mathbb{Q}_{n,\theta}\left(\log \mathbb{H}_{n,\theta}\right) \geq \mathbb{Q}_{n,\theta}\left(\log \mathbb{H}_{n,\theta'}\right)$$

Also observe that

$$\log \mathbb{H}_{n,\theta} = -\log \mathcal{Z}_{n,\theta} + \mathbb{L}_{n,\theta}$$

with the additive functional

$$\mathbb{L}_{n,\theta}(x_0, \ldots, x_n) := \sum_{p=0}^{n} \log H_{\theta,p}(x_{p-1}, x_p)$$

One concludes that for any pair of parameters (θ, θ')

$$\mathbb{Q}_{n,\theta}\left(\log H_{\theta',n}\right) = -\log \mathcal{Z}_{n,\theta'} + \mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta'}\right) \leq -\log \mathcal{Z}_{n,\theta} + \mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta}\right)$$

and therefore

$$\mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta}\right) - \mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta'}\right) \geq \log\left(\mathcal{Z}_{n,\theta}/\mathcal{Z}_{n,\theta'}\right)$$

In other words, we have

$$\mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta}\right) \leq \mathbb{Q}_{n,\theta}\left(\mathbb{L}_{n,\theta'}\right) \Rightarrow \mathcal{Z}_{n,\theta} \leq \mathcal{Z}_{n,\theta'}$$

We denote by $(\theta_k)_{k\geq 0}$ a sequence of parameters starting at some state θ_0 and defined by the following recursion

$$\theta_k := \max_{\theta'} \mathbb{Q}_{n,\theta_{k-1}}\left(\mathbb{L}_{n,\theta'}\right) \implies \mathcal{Z}_{n,\theta_k} \geq \mathcal{Z}_{n,\theta_{k-1}}$$

If $\mathbb{Q}_{n,\theta}$ is in the exponential family, then the maximization step is usually straightforward. More precisely, there exists a collection of functions $(f_{n,\theta}^{(i)})_{i\in\mathcal{I}}$, indexed by some finite set $\mathcal{I}$, on $\boldsymbol{E_n}$ and some $F_{n,\theta} : \mathbb{R}^{\mathcal{I}} \mapsto \mathbb{R}^d$ such that

$$\theta_k = F_{n,\theta_{k-1}}\left(\left[\mathbb{Q}_{n,\theta_{k-1}}\left(f_{n,\theta_{k-1}}^{(i)}\right)\right]_{i\in\mathcal{I}}\right) \tag{11.28}$$

The set of functions $(f_n^{(i,\theta)})_{i\in\mathcal{I}}$ is sometimes referred to as sufficient statistics in the literature.

As shown in [204], one way to approximate the recursive Equation (11.28) consists of replacing the measures $\mathbb{Q}_{n,\theta_{n-1}}$ by the backward particle measures $\mathbb{Q}^N_{n,\theta_{n-1}}$, or by the genealogical tree occupation measures $\boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{n},\boldsymbol{\theta_{n-1}}}$ defined in (10.6). The corresponding mean field IPS approximation models are defined by

$$\theta_k^N = F_{n,\theta_{k-1}^N}\left(\left[\mathbb{Q}^N_{n,\theta_{k-1}}\left(f_{n,\theta_{k-1}^N}^{(i)}\right)\right]_{i\in\mathcal{I}}\right)$$

or by

$$\theta_k^N = F_{n,\theta_{k-1}^N}\left(\left[\boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{n},\boldsymbol{\theta_{n-1}}}\left(f_{n,\theta_{k-1}^N}^{(i)}\right)\right]_{i\in\mathcal{I}}\right)$$

For a more thorough discussion on these stochastic algorithms and their convergence analysis, we refer the reader to the article [204].

11.3.6 Stochastic gradient algorithms

We come back to parametrized models presented in Section 11.3.5. One alternative way of computing the maximum value of the mapping $\theta \mapsto \mathcal{Z}_{n,\theta}$ is to introduce a more conventional gradient type steepest descent model

$$\theta_k = \theta_{k-1} + \tau_k \, \nabla \log \mathcal{Z}_{n,\theta_{k-1}}$$

with a positive real sequence of parameters τ_n such that $\sum_k \tau_k = \infty$ and $\sum_k \tau_k^2 < \infty$. Using (10.50), we have

$$\nabla \log \mathcal{Z}_{n,\theta} = \mathbb{Q}_{n,\theta}(\Lambda_{n,\theta})$$

with the additive functional

$$\Lambda_{n,\theta}(x_0, \ldots, x_n) := \sum_{p=0}^{n} \nabla \log H_{\theta,p}(x_{p-1}, x_p)$$

We can approximate these equations using the following equations

$$\theta_k^N = \theta_{k-1}^N + \tau_k \, \nabla_N \log \mathcal{Z}_{n,\theta_{k-1}} = \theta_{k-1}^N + \tau_k \, \mathbb{Q}^N_{n,\theta_{k-1}}(\Lambda_{n,\theta^N_{k-1}})$$

with the particle derivatives associated with the backward particle measures $\mathbb{Q}^N_{n,\theta_{n-1}}$. We can alternatively use the recursion

$$\theta_k^N = \theta_{k-1}^N + \tau_k \, \nabla_N \log \mathcal{Z}_{n,\theta_{k-1}} = \theta_{k-1}^N + \tau_k \, \boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{n},\boldsymbol{\theta_{k-1}}}(\Lambda_{n,\theta^N_{k-1}})$$

with the genealogical tree occupation measures $\boldsymbol{\eta}^{\boldsymbol{N}}_{\boldsymbol{n},\boldsymbol{\theta_{n-1}}}$ defined in (10.6). For a more thorough discussion on these particle steepest descent algorithms and their connections with filter derivative models, we refer the reader to the article [205]. The convergence analysis of these models can be developed using the stochastic analysis techniques presented in the textbook [39].

11.4 Risk analysis and rare event simulation

The analysis of rare events arises in various scientific areas including physics, biology, engineering science, and financial mathematics.

For instance, in nuclear physics, we might be interested in computing the probability that some radiation escapes from some containment before being absorbed by some obstacle. In biology, these rare events may be related to extinction probabilities of some population evolution model. In engineering sciences, these critical events are often related to a catastrophic failure, such as a buffer overflows in communication networks. Finally, in financial mathematics, they arise in the analysis of portfolio credit risk models. In this context, the critical events represent ruin processes, or credit payment default probabilities. Importance sampling techniques are perhaps one of the most widely used alternative to crude Monte Carlo simulation of unlikely events. The idea is to generate samples from a different judiciously chosen distribution, rather than from the distribution of interest. These statistical techniques have two main drawbacks. Very often, the twisted distribution cannot be chosen as we would like, since we need to have a dedicated technique to sample random variables w.r.t. these measures. On the other hand, these importance sampling techniques are intrusive in the sense that we need to twist the reference random process, so that to produce unphysical trajectories. In Section 11.4.1, we present a nonintrusive mean field IPS technique for the simulation of importance sampling distributions without altering the nature of the reference process. Further details on these models, including applications in fiber optics communication and financial risk analysis, can also be found in a couple of articles [104, 180, 181].

Section 11.4.2 is dedicated to mean field multilevel simulation. These techniques are often termed multilevel splitting particle methods or sequential Monte Carlo samplers in the literature on rare event simulation, Further detail on these IPS models can be found in the review article [193], as well as in the series of articles [132, 133, 363].

The final section, Section 11.4.3, is concerned with the mean field IPS computation of Dirichlet problems with boundary conditions. These problems arise in a variety of application areas of physics, including fluid mechnanics and plasma dynamics, as well as in optics and traffic engineering. For a detailed discussion on these problems in the context of elliptic-hypebolic equations of Keldysh type we refer the reader to the monograph [473].

11.4.1 Importance sampling and twisted measures

Computing the probability of some events of the form $\{V_n(X_n) \geq a\}$, for some energy like function V_n and some threshold a, is often performed using the importance sampling distribution of the state variable X_n with some multiplicative Boltzmann weight function $\exp(\beta V_n(X_n))$, associated with some inverse temperature parameter β. These twisted measures can be described by a Feynman-Kac model in transition space by setting

$$G_n(X_{n-1}, X_n) = \exp\{\beta[V_n(X_n) - V_{n-1}(X_{n-1})]\}$$

For instance, it is easily checked that

$$\mathbb{P}(V_n(X_n) \geq a) \quad = \quad \mathbb{E}\left(\mathbf{f_n}(\mathbf{X}_n) \prod_{0 \leq p < n} G_p(\mathbf{X}_p)\right)$$

with the function $f_n(\mathbf{X}_n) = 1_{V_n(X_n) \geq a}\, e^{-\beta V_n(X_n)}$, and the potential function and the reference Markov chain

$$\mathbf{X}_n = (X_n, X_{n+1}) \quad \text{and} \quad G_n(\mathbf{X}_n) = \exp\{\beta(V_{n+1}(X_{n+1}) - V_n(X_n))\}$$

We let $\mathbb{Q}_n$ be the Feynman-Kac model (4.17) associated with the reference Markov chain $\mathbf{X}_n$ and the potential function G_n. In the same vein, we have the Feynman-Kac formulae

$$\mathbb{E}\left(\varphi_n(X_0,\ldots,X_n) \mid V_n(X_n) \geq a\right) = \mathbb{Q}_n(F_{n,\varphi_n})/\mathbb{Q}_n(F_{n,1})$$

with the function $F_{n,\varphi_n}(X_0,\ldots,X_n) = \varphi_n(X_0,\ldots,X_n)\ 1_{V_n(X_n)\geq a}\ e^{-\beta V_n(X_n)}$. The mean field IPS simulation of these Feynman-Kac distributions is defined in Section 9.3.

11.4.2 Multilevel splitting simulation

We consider some Markov chain $(X'_n)_{n\geq 0}$ taking values in some finite state space E'. We assume that the chain X'_n starts in some given subset $X'_0 \in A \subset E'$ with a given distribution ν_0. We also let (B,C) be a pair of subsets (B,C) such that $A\cap C = \emptyset = B\cap C$. We also assume that the triplet (A,B,C) is chosen so that for any initial state $x\in A$ the chain X'_n hits one of the sets B, or C in finite time.

We let T_A be the entrance time of X' into a given subset A. One would like to estimate the probability that the chain hits B before C

$$\mathbb{P}(T_{B\cup C} < T_C) = \mathbb{P}(X'_{T_{B\cup C}} \in B) = \mathbb{E}(1_B(X'_{T_{B\cup C}}))$$

and the law of the excursions given the fact that it reached B before C

$$\mathrm{Law}(X'_t\ ;\ 0\leq t\leq T_{B\cup C} \mid T_{B\cup C} < T_C)$$

Of course we have implicitly assumed that $\mathbb{P}(T_{B\cup C} < T_C) > 0$ so that the conditional distributions are well defined. During its excursion from A to B, the process eventually visits a decreasing sequence of level sets $(B_n)_{n=0,\ldots,m}$

$$A = B_0 \supset B_1 \supset \ldots \supset B_m = B \tag{11.29}$$

This decomposition reflects the successive gateways the stochastic process needs to cross before entering into the relevant rare event.

To simplify the presentation, we slightly abuse the notation, and we write T_n instead of $T_{B_n\cup C}$. In this simplified notation, to capture the behavior of X between the different levels we introduce the excursion-valued Markov chain

$$X_n = (T_n, (X'_t\ ;\ T_{n-1}\leq t\leq T_n)) \in E = \cup_{p\leq q}(\{q\}\times (E')^{(q-p+1)}) \tag{11.30}$$

Under our assumptions, these entrance times are finite and

$$(T_{B\cup C} < T_C) = (T_m < T_C) = \bigcap_{1\leq p\leq m} (T_p < T_C)$$

To check whether or not the n-th excursion has reached the desired n-th level, we consider the potential functions on E defined for each $n\in\{0,\ldots,m\}$ and $x = (x_q)_{p\leq q\leq r} \in (E')^{(r-p+1)}$, by $G_n(r,x) = 1_{B_n}(x_r)$. In this notation we have for each $n\leq m$

$$(T_n < T_C) = \textstyle\bigcap_{1\leq p\leq n}(T_p < T_C) = \bigcap_{1\leq p\leq n}(G_p(X_p) = 1)$$

$$(X_0,\ldots,X_n)$$

$$= ((0,X'_0),(T_1,(X'_t\ ;\ 0\leq t\leq T_1)),\ldots,(T_n,(X'_t\ ;\ T_{n-1}\leq t\leq T_n)))$$

If we write $[X'_t\ ;\ 0\leq t\leq T_n]$ instead of $(X_0,\ldots,X_n)$, the sequence of excursions of X' between the levels, then for any $n\leq m$ and any function f_n on the product space E^{n+1} we have the Feynman-Kac formulae

$$\mathbb{E}_{\nu_0}\left(f_n(X_0,\ldots,X_n)\ \prod_{p=1}^{n} G_p(X_p)\right) = \mathbb{E}_{\nu_0}\left(f_n([X'_t\ ;\ 0\leq t\leq T_n])\ 1_{T_n<T_C}\right) \tag{11.31}$$

The mean field IPS simulation of these Feynman-Kac distributions is defined in Section 9.3.

11.4.3 Dirichlet problems with hard boundary conditions

We consider the same excursion model discussed in Section 11.4.2 but we replace the potential function by the function G_n on E defined for any $n \in \{0, \ldots, m\}$, $0 \leq p \leq r$, and $x = (x_q)_{p \leq q \leq r} \in (E')^{(r-p+1)}$ by

$$G_n(r, x) = 1_{B_n}(x_r) \prod_{p<q\leq r} G'(x_q) \tag{11.32}$$

with some nonnegative functions G' on the finite set E'. In this situation, the r.h.s. expectation in (11.31) is given by

$$\mathbb{E}_{\nu_0}\left(f_n([X'_t\,;\, 0 \leq t \leq T_n])\; 1_{B_n}(X'_{T_n}) \prod_{p=1}^{T_n} G'(X'_p) \right)$$

We recall that these Feynman-Kac formulae can be computed using mean field IPS simulation. Next, we examine the excursion-valued models (11.29) when $A = E' - C$, and $(B_n \cap C) = \emptyset$. For $\nu_0 = \delta_x$, with $x \in A$, $n = m$, and any function f on E', the above expectations are given by the following function

$$h(x) = \mathbb{E}_x\left(f(X'_T) 1_B(X'_T) \prod_{p=1}^{T} G'(X'_p) \right)$$

with the first time $T(= T_m)$ the process X' hits the domain $D := (B \cup C)$. Using the fact that $x \in D \Rightarrow T = 0$, we extend the function h on C by setting $h(x) = f(x)1_B(x)$. By a conditioning argument, for any $x \notin D$, we have

$$h(x) = \mathbb{E}_x\left(G'(X'_1)\; \mathbb{E}_{X'_1}\left(f(X'_T)1_B(X'_T) \prod_{p=1}^{T} G'_p(X'_p) \right) \right) = \mathbb{E}_x\left(G'(X'_1)h(X'_1) \right)$$

From the above discussion, if $M'(x, y)$ is the Markov transition of the chain X', then we see that the function h satisfies the following Dirichlet problem with hard boundary conditions

$$\left\{ \begin{array}{rcll} M'(G'h)(x) & = & h(x) & \text{for} \quad x \notin D \\ h(x) & = & f(x)1_B(x) & \text{for} \quad x \in D \end{array} \right.$$

For a more thorough discussion on the Dirichlet problem for more general models, we refer the reader to the book [172].

Chapter 12

Nonlinear evolutions of intensity measures

12.1 Intensity of spatial branching processes

The Feynman-Kac models presented in section 9.1.4 were defined in terms of Markov chain distributions, weighted by some potential functions. This description is particularly useful to model conditional distributions of Markov chains w.r.t. a collection of conditioning events. In this section, we present a natural and alternative interpretation of these models in terms of spatial branching processes. We also extend the Feynman-Kac methodology to branching models, equipped with spontaneous birth rates.

For a more detailed discussion on spatial branching processes, and their connections with Feynman-Kac models, we also refer the reader to [21, 335, 358]. Section 12.3 and the more recent studies [110, 172] also provide applications of these models to multiple object nonlinear filtering problems.

12.1.1 Spatial branching processes

Assume that, at a given time n, there are N_n individuals $(X_n^i)_{1\leq i\leq N_n}$, taking values in some measurable state space E_n, enlarged with an auxiliary cemetery point c. As usual, we extend the measures γ_n on E_n and the bounded measurable functions f_n on E_n by setting $\gamma_n(\{c\}) = 0$ and $f_n(c) = 0$.

We emphasize that the state space E_n depends on the problem at hand. It may vary with the time parameter, and it can include all the characteristics of an individual, such as its type, its kinetic parameters as well as its complete path from the origin.

Each individual X_n^i has a survival probability, say $e_n(X_n^i) \in [0, 1]$. When it dies, it goes instantly to the cemetery point c. We also use the convention $e_n(c) = 0$, so that a dead particle can only stay in the cemetery state.

Survival particles give birth to a random strictly positive number of individuals $h_n^i(X_n^i)$ where $\left(h_n^i(x_n)\right)_{1\leq i\leq N_n}$ stands for a collection of independent random variables such that

$$\mathbb{E}\left(h_n^i(x_n)\right) = H_n(x_n)$$

for any $x_n \in E_n$, where H_n is a given collection of bounded nonnegative functions.

Notice that $H_n(x_n) \geq 1$ for any $x_n \in E_n$, since $h_n^i(x_n) \geq 1$. This branching transition is sometimes called spawning in signal processing and multi-target tracking literature.

After this branching transition, the system consists of a random number $\widehat{N}_n$ of individuals $(\widehat{X}_n^i)_{1\leq i\leq \widehat{N}_n}$. Each of them evolves randomly, and independently, from state space to the next

$$\widehat{X}_n^i = x_n \in E_n \rightsquigarrow X_{n+1}^i = x_{n+1} \in E_{n+1}$$

according to a Markov transition $M_{n+1}(x_n, dx_{n+1})$ from E_n into E_{n+1}. Here again, we use the convention $M_{n+1}(c, \{c\}) = 1$, so that any dead particle remains in the cemetery state.

At the same time, an independent collection of new individuals is added to the current configuration.

We further assume that this additional spatial point process is modeled by a spatial Poisson process, with a prescribed intensity measure μ_{n+1} on E_{n+1}. It is used to model new particles entering the state space.

At the end of this transition, we obtain

$$N_{n+1} = \widehat{N}_n + N'_{n+1}$$

individuals $(X^i_{n+1})_{1\leq i\leq N_{n+1}}$, where N'_{n+1} is a Poisson random variable with parameters given by the total mass $\mu_{n+1}(1)$ of the positive measure μ_{n+1}, and $(X^{\widehat{N}_n+i}_{n+1})_{1\leq i\leq N'_{n+1}}$ are independent and identically distributed random variables with common distribution

$$\overline{\mu}_{n+1} = \mu_{n+1}/\mu_{n+1}(1) \quad \text{where} \quad \mu_{n+1}(1) := \int_{E_{n+1}} \mu_{n+1}(dx) = \mathbb{E}\left(N'_{n+1}\right) \tag{12.1}$$

For a more thorough discussion on spatial Poisson point processes, we refer the reader to Section 12.3.3.

We end this section with some definitions, some conventions, and a few regularity conditions that will implicitly be assumed in the further development of this chapter.

Firstly, to simplify the presentation, we shall further assume that the initial configuration of the spatial branching process $(X^i_0)_{1\leq i\leq N_0}$ is given by a spatial Poisson process, with a prescribed intensity measure μ_0 on E_0.

We denote by G_n the potential functions defined by

$$x_n \in E_n \mapsto G_n(x_n) = e_n(x_n)H_n(x_n)$$

To avoid unnecessary technical details, we further assume that the potential functions G_n are chosen so that for any $x \in E_n$

$$0 < g_{n,-} \leq G_n(x) \leq g_{n,+} < \infty \tag{12.2}$$

for any time parameter $n \geq 0$.

Note that this assumption is satisfied in most realistic spatial branching scenarios. Indeed, as $H_n(x) \geq 1$, the condition $g_{n,-} \leq G_n(x)$ essentially states that there exists $e_{n,-} > 0$ such that $e_n(x) \geq e_{n,-}$ for any $x \in E_n$. Loosely speaking, this condition ensures that every particle always has a small chance to survive.

On the other hand, the condition $G_n(x) \leq g_{n,+}$ states that there exists $H_{n,+} < \infty$ such that $H_n(x) \leq H_{n,+}$ for any $x \in E_n$. Loosely speaking, this condition allows controlling the total size of the branching process by some rather crude, but bounded finite constants.

In the unlikely scenario where (12.2) is not satisfied, the forthcoming analysis can be extended to more general models with general nonnegative potential functions, using the techniques developed in Section 4.4 in the monograph [172].

12.1.2 Intensity distribution flows

In this section we discuss the evolution equation of the intensity measures associated with the spatial branching model presented in Section 12.1.1.

Definition 12.1.1 *We denote by $\mathcal{X}_n$ the occupation measure of the branching particle model*

$$\mathcal{X}_n := \sum_{i=1}^{N_n} \delta_{X^i_n}$$

The intensity measure γ_n associated with this point process is given for any bounded measurable function f_n on $E_{n,c} = E_n \cup \{c\}$ by the following formula:

$$\gamma_n(f_n) := \mathbb{E}\left(\mathcal{X}_n(f)\right) \qquad with \quad \mathcal{X}_n(f_n) := \int \mathcal{X}_n(dx_n)\ f_n(x_n)$$

To simplify the presentation, we suppose that the initial configuration of the particles is a spatial Poisson process with intensity measure μ_0 on the state space E_0.

Given the construction defined in Section 12.1.1, it follows almost straightforwardly that the intensity measures γ_n on E_n satisfy the following recursive equation.

Lemma 12.1.2 *For any $n \geq 0$, we have*

$$\gamma_{n+1} = \gamma_n Q_{n+1} + \mu_{n+1} \tag{12.3}$$

with the initial condition $\gamma_0 = \mu_0$. In the above displayed formulae, μ_{n+1} is the intensity measure of the spatial point process associated to the birth of new individuals at time $n+1$, while Q_{n+1} is the bounded and positive integral operator from E_n into E_{n+1}

$$Q_{n+1}(x_n, dx_{n+1}) := G_n(x_n)\ M_{n+1}(x_n, dx_{n+1}) \tag{12.4}$$

Proof:
For any bounded measurable function f on $E_{n+1} \cup \{c\}$, we have

$$\gamma_{n+1}(f) = \mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) + \mathbb{E}\left(\sum_{i=\widehat{N}_n+1}^{\widehat{N}_n+N'_{n+1}} f\left(X_{n+1}^i\right)\right)$$

Thanks to the Poisson assumption, we have

$$\mathbb{E}\left(\sum_{i=\widehat{N}_n+1}^{\widehat{N}_n+N'_{n+1}} f\left(X_{n+1}^i\right)\right) = \mu_{n+1}(1)\,\overline{\mu}_{n+1}(f) = \mu_{n+1}(f)$$

with the normalized measures $\overline{\mu}_{n+1}$ defined in (12.1).

We let $\widehat{\mathcal{F}}_n$ be the σ-field generated by $(\widehat{X}_n^i)_{1\leq i\leq \widehat{N}_n}$ and $\mathcal{F}_n$ the σ-field generated by $\left(X_n^i\right)_{1\leq i\leq N_n}$. In this notation, we have that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) &= \mathbb{E}\left(\mathbb{E}\left[\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right) \,\middle|\, \widehat{\mathcal{F}}_n\right]\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[\sum_{i=1}^{\widehat{N}_n} M_{n+1}(f)\left(\widehat{X}_n^i\right) \mid \mathcal{F}_n\right]\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{N_n} e_n\left(X_n^i\right) H_n(X_n^i) M_{n+1}(f)\left(X_n^i\right)\right)
\end{aligned}$$

from which we conclude that

$$\mathbb{E}\left(\sum_{i=1}^{\widehat{N}_n} f\left(X_{n+1}^i\right)\right) = \gamma_n\left(e_n H_n M_{n+1}(f)\right) = \gamma_n\left(G_n M_{n+1}(f)\right)$$

and therefore

$$\gamma_{n+1}(f) = \gamma_n(Q_{n+1}(f)) + \mu_{n+1}(f)$$

This ends the proof of the lemma. ∎

The flow of intensity measures γ_n is clearly more complex that the Feynman-Kac distribution flows (9.7) discussed in Section 9.1.4. Thus, we typically do not expect to find any closed-form expression to solve these equations. A natural way to approximate them numerically is to use a mean field particle interpretation of the associated sequence of normalized probability distributions.

Definition 12.1.3 *The normalized probability distributions associated with the intensity distributions γ_n are the probability measures $\eta_n \in \mathcal{P}(E_n)$ defined for any $f_n \in \mathcal{B}_b(E_n)$ by*

$$\eta_n(f_n) := \gamma_n(f_n)/\gamma_n(1)$$

We end this section, with a couple of remarks. When $\mu_n = 0$, the distributions (γ_n, η_n) coincide with the Feynman-Kac models (9.7) discussed in Section 9.1.4. On the other hand, we notice that

$$\mathbb{E}(N_n) = \gamma_n(1) \quad \text{and} \quad \eta_n(f_n) = \frac{1}{\mathbb{E}(N_n)}\,\mathbb{E}\left(\sum_{i=1}^{N_n} f_n(X_n^i)\right)$$

12.1.3 Nonlinear evolution equations

In this section, we discuss the evolution equations of the normalized probability measures η_n, introduced in Definition 12.1.3. In contrast with conventional Feynman-Kac models, these extended models are expressed in terms of updating-prediction transitions that depend on the total mass $\gamma_n(1)$ of the intensity measures γ_n.

To describe with some conciseness these models, we need another round of notation. In subsequent pages of this section, we identify the measures γ_n with a couple of parameters $(\gamma_n(1), \eta_n)$. The first component $\gamma_n(1)$ represents the total mass of γ_n, and η_n the normalized probability measure.

Definition 12.1.4 *We consider the collection of Markov transitions $M_{n+1,(m,\eta)}$ indexed by the parameters $m \in \mathbb{R}_+$ and the probability measures $\eta \in \mathcal{P}(E_n)$ defined by*

$$M_{n+1,(m,\eta)}(x, dy) := \alpha_n(m,\eta)\, M_{n+1}(x, dy) + (1 - \alpha_n(m,\eta))\ \overline{\mu}_{n+1}(dy) \tag{12.5}$$

with the collection of $[0,1]$-valued functions

$$\alpha_n\ :\ (m,\eta) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))\ \mapsto\ \alpha_n(m,\eta) = \frac{m\eta(G_n)}{m\eta(G_n) + \mu_{n+1}(1)}$$

Definition 12.1.5 *We let Λ_{n+1} be the mapping from $\mathbb{R}_+ \times \mathcal{P}(E_n)$ into $\mathbb{R}_+ \times \mathcal{P}(E_{n+1})$ given by*

$$\Lambda_{n+1}(m,\eta) = \left(\Lambda^1_{n+1}(m,\eta), \Lambda^2_{n+1}(m,\eta)\right) \tag{12.6}$$

with the pair of transformations:

$$\begin{aligned} \Lambda^1_{n+1}(m,\eta) &= m\ \eta(G_n) + \mu_{n+1}(1) \\ \Lambda^2_{n+1}(m,\eta) &= \Psi_{G_n}(\eta) M_{n+1,(m,\eta)} \end{aligned}$$

In the above display, Ψ_{G_n} stands for the Boltzmann-Gibbs transformation associated with a potential function G_n, defined in (0.2).

The semigroup of the flow γ_n, or equivalently $(\gamma_n(1), \eta_n)$, is now expressed in terms of the mathematical objects defined above.

Proposition 12.1.6 *For any $n \geq 0$, we have the evolution equations*

$$(\gamma_n(1), \eta_n) = \Lambda_n(\gamma_{n-1}(1), \eta_{n-1}) \tag{12.7}$$

Proof:
Observe that for any function $f \in \mathcal{B}(E_{n+1})$, we have that

$$\eta_{n+1}(f) = \frac{\gamma_n(G_n M_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n(G_n) + \mu_{n+1}(1)} = \frac{\gamma_n(1)\ \eta_n(G_n M_{n+1}(f)) + \mu_{n+1}(f)}{\gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1)}$$

from which we find that

$$\eta_{n+1} = \alpha_n\left(\gamma_n(1), \eta_n\right)\ \Psi_{G_n}(\eta_n) M_{n+1} + \left(1 - \alpha_n\left(\gamma_n(1), \eta_n\right)\right)\ \overline{\mu}_{n+1}$$

From these observations, we prove (12.7). This ends the proof of the proposition. ∎

12.1.4 McKean interpretations

In this section, we design a McKean interpretation of the measure valued process $(\gamma_n(1), \eta_n) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))$ introduced in Section 12.1.3.

In Proposition 12.1.6, we have shown that the evolution Equation (12.7) of the sequence of probability measures $\eta_n \rightsquigarrow \eta_{n+1}$ is a combination of an updating type transition $\eta_n \rightsquigarrow \Psi_{G_n}(\eta_n)$, and an integral transformation w.r.t. a Markov transition $M_{n+1,(\gamma_n(1),\eta_n)}$ that depends on the current total mass $\gamma_n(1)$, as well as on the current probability distribution η_n.

The integral operator $M_{n+1,(\gamma_n(1),\eta_n)}$ defined in (12.5) is a mixture of the Markov transition M_{n+1} and the spontaneous birth normalized measure $\overline{\mu}_{n+1}$. Notice that for null spontaneous birth measures, this Markov transition reduces to the one of the free exploration of the particles; that is, we have that

$$\mu_{n+1} = 0 \implies M_{n+1,(\gamma_n(1),\eta_n)} = M_{n+1}$$

We let S_{n,η_n} be any Markov transition from E_n into itself satisfying the following compatibility condition

$$\Psi_{G_n}(\eta_n) = \eta_n S_{n,\eta_n}$$

Several examples of transitions S_{n,η_n} are discussed in (8).

By construction, we have the recursive formula

$$\eta_{n+1} = \eta_n K_{n+1,(\gamma_n(1),\eta_n)} \tag{12.8}$$

with the Markov transitions

$$K_{n+1,(\gamma_n(1),\eta_n)} = S_{n,\eta_n} M_{n+1,(\gamma_n(1),\eta_n)}$$

The sequence of probability distributions η_n can be interpreted as the distributions of the random states $\overline{X}_n$ of a process defined, conditional upon $(\gamma_n(1), \eta_n)$, by the elementary transitions

$$\mathbb{P}\left(\overline{X}_{n+1} \in dx \mid \overline{X}_n\right) = K_{n,(\gamma_n(1),\eta_n)}\left(\overline{X}_n, dx\right) \quad \text{with} \quad \eta_n = \text{Law}(\overline{X}_n)$$

and the auxiliary total mass evolution equation

$$\gamma_{n+1}(1) = \gamma_n(1)\ \eta_n(G_n) + \mu_{n+1}(1) \tag{12.9}$$

The transport formula presented in (12.8) provides a natural interpretation of the probability distributions η_n as the laws of a process $\overline{X}_n$ whose elementary transitions $\overline{X}_n \rightsquigarrow \overline{X}_{n+1}$ depend on the distribution $\eta_n = \text{Law}(\overline{X}_n)$ as well as on the current mass $\gamma_n(1)$.

In contrast to the more traditional McKean type nonlinear Markov chains discussed in Section 9, the dependency on the mass process induces a dependency on the whole sequence of measures η_p, from the origin $p = 0$ up to the current time $p = n$.

12.1.5 Mean field particle interpretation

In this section, we design a mean field interpretation of the McKean models developed in Section 12.1.4. From now on, we will always assume that the mappings

$$\left(m, \left(x^i\right)_{1\leq i\leq N}\right) \in \left(\mathbb{R}_+ \times E_n^N\right) \mapsto K_{n+1,\left(m,\frac{1}{N}\sum_{j=1}^N \delta_{x^j}\right)}\left(x^i, A_{n+1}\right)$$

are measurable w.r.t. the product σ-fields on $(\mathbb{R}_+ \times E_n^N)$, for any $n \geq 0$, $N \geq 1$, and $1 \leq i \leq N$, and any measurable subset $A_{n+1} \subset E_{n+1}$.

In this situation, the mean field particle interpretation of (12.8) and (12.9) is the Markov chain

$$(\gamma_n^N(1), \xi_n) \in (\mathbb{R}_+ \times E_n^N) \qquad \text{with} \quad \xi_n = \left(\xi_n^i\right)_{1\leq i\leq N} \in E_n^N$$

and with elementary transitions

$$\left\{\begin{array}{rcl} \mathbb{P}\left(\xi_{n+1} \in dx \;\middle|\; (\gamma_n^N(1), \xi_n)\;\right) &=& \prod_{i=1}^N\; K_{n+1,(\gamma_n^N(1),\eta_n^N)}(\xi_n^i, dx^i) \\ \gamma_{n+1}^N(1) &=& \gamma_n^N(1)\;\eta_n^N(G_n) + \mu_{n+1}(1) \end{array}\right. \tag{12.10}$$

with the infinitesimal neighborhood $dx = dx^1 \times \ldots \times dx^N$ of a point $x = (x^1, \ldots, x^N) \in E_{n+1}^N$. In the above displayed formula, $\left(\gamma_n^N, \eta_n^N\right)$ stand for the couple of occupation measures defined for any $f_n \in \mathcal{B}_b(E_n)$ by

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \quad \text{and} \quad \gamma_n^N(f_n) := \gamma_n^N(1)\;\eta_n^N(f_n)$$

The initial system ξ_0 consists of N independent and identically distributed random variables with common law η_0, and we assume that the initial mass $\gamma_0^N(1) = \gamma_0(1) = \mu_0(1)$ is explicitly known. In this connection, we mention that the particle total mass model is also given by the following formula

$$\gamma_n^N(1) = \sum_{p=0}^n\; \mu_p(1) \prod_{p\leq q<n} \eta_q^N(G_q)$$

By definition of the two step McKean transitions (12.8), the mean particle evolution described by (12.10) is a "simple" combination of a selection and a mutation genetic type transition

$$\xi_n \in E_n^N \quad \rightsquigarrow \quad \widehat{\xi}_n = (\widehat{\xi}_n^i)_{1\leq i\leq N} \in E_n^N \quad \rightsquigarrow \quad \xi_{n+1} \in E_{n+1}^N$$

During the selection transitions $\xi_n \rightsquigarrow \widehat{\xi}_n$, each particle $\xi_n^i \rightsquigarrow \widehat{\xi}_n^i$ evolves according to the selection type transition $S_{n,\eta_n^N}(\xi_n^i, dx)$. During the mutation stage, each of the selected particles $\widehat{\xi}_n^i \rightsquigarrow \xi_{n+1}^i$ evolves according to the transition

$$M_{n+1,(\gamma_n^N(1),\eta_n^N)}(x, dy) := \alpha_n\left(\gamma_n^N(1), \eta_n^N\right) M_{n+1}(x, dy) + \left(1 - \alpha_n\left(\gamma_n^N(1), \eta_n^N\right)\right)\; \overline{\mu}_{n+1}(dy)$$

12.2 Nonlinear equations of positive measures

12.2.1 Measure valued evolution equations

We consider a general class of measure-valued processes $\gamma_n \in \mathcal{M}_+(E_n)$ defined by the following nonlinear equations

$$\gamma_n = \Xi_n(\gamma_{n-1}) := \gamma_{n-1} Q_{n,\gamma_{n-1}} \tag{12.11}$$

with initial measure $\gamma_0 \in \mathcal{M}_+(E_0)$. In the above display, $Q_{n,\gamma}$ stands for a collection of positive and bounded integral operators from E_{n-1} into E_n, indexed by the time parameter $n \geq 1$, and the set of measures $\gamma \in \mathcal{M}_+(E_{n-1})$.

One natural way to solve the nonlinear integral Equation (12.11) is to use a judicious probabilistic interpretation of the normalized distributions flow given for any $f_n \in \mathcal{B}_b(E_n)$ by

$$\eta_n(f_n) := \gamma_n(f_n)/\gamma_n(1)$$

To describe with some conciseness these stochastic models, it is important to observe that the pair process $(\gamma_n(1), \eta_n) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))$ satisfies an evolution equation of the following form

$$(\gamma_n(1), \eta_n) = \Lambda_n(\gamma_{n-1}(1), \eta_{n-1}) \tag{12.12}$$

for some mapping

$$\Lambda_n \ : \ (m, \eta) \in (\mathbb{R}_+ \times \mathcal{P}(E_{n-1})) \mapsto \Lambda_n(m, \eta) \in (\mathbb{R}_+ \times \mathcal{P}(E_n))$$

We also denote by $\left(\Lambda_n^1, \Lambda_n^2\right)$, the first, and second, component mappings of the one step transformation Λ_n given by

$$\Lambda_n^1 \ : \ (\mathbb{R}_+ \times \mathcal{P}(E_n)) \ \to \ \mathbb{R}_+ \quad \text{and} \quad \Lambda_n^2 \ : \ (\mathbb{R}_+ \times \mathcal{P}(E_n)) \ \to \ \mathcal{P}(E_n)$$

By construction, we notice that the total mass process can be computed using the recursive formula

$$\gamma_{n+1}(1) = \gamma_n(G_{n,\gamma_n}) = \eta_n(G_{n,\gamma_n})\ \gamma_n(1) \quad \text{with} \quad G_{n,\gamma_n} := Q_{n+1,\gamma_n}(1) \tag{12.13}$$

On the other hand, for any $f \in \mathcal{B}_b(E_{n+1})$ we have that

$$\eta_{n+1} = \Psi_{G_{n,\gamma_n}}(\eta_n)\, M_{n+1,\gamma_n} \quad \text{with} \quad M_{n+1,\gamma_n}(f) := \frac{Q_{n+1,\gamma_n}(f)}{Q_{n+1,\gamma_n}(1)}$$

and the Boltzmann-Gibbs transformation Ψ_G associated with the potential function $G = G_{n,\gamma_n}$ and defined in (0.2). This implies that

$$\Lambda_{n+1}^1(m, \eta) = m\ \eta(G_{n,m\eta}) \quad \text{and} \quad \Lambda_{n+1}^2(m, \eta) = \Psi_{G_{n,m\eta}}(\eta_n)\, M_{n+1,m\eta} \tag{12.14}$$

We end this section with some comments on the applications and the stability analysis of these rather abstract models. We also provide some reference pointers to the sections of the book discussing in more detail these questions.

Illustrations in the context of multiple target tracking problems are presented in Section 12.3.2.

12.2.2 Mean field particle models

The mean field particle model associated with the Equation (12.12) relies on the fact that the one step mappings Λ_{n+1}^2 given in (12.14) can be rewritten in terms of the nonlinear Markov transport equations

$$\Psi_{G_{n,\gamma}}(\eta)\, M_{n,\gamma} = \eta K_{n+1,\gamma} \quad \text{with} \quad K_{n+1,\gamma} := S_{n,\gamma} M_{n+1,\gamma} \tag{12.15}$$

for any $\gamma = m\eta \in \mathcal{M}_+(E_n)$. In the above displayed formula, $S_{n,\gamma}$ stands for any collection of Markov transitions, from E_n into itself, and index by $\gamma \in \mathcal{M}_+(E_n)$, satisfying the following compatibility condition

$$\Psi_{G_{n,\gamma}}(\eta) = \eta S_{n,\gamma}$$

Several examples of transitions $S_{n,m\eta}$ are discussed in (8).

These models provide a natural interpretation of the distribution laws η_n as the laws of a nonlinear Markov chain $\overline{X}_n$ whose elementary transitions $\overline{X}_n \rightsquigarrow \overline{X}_{n+1}$ depend on the distribution $\eta_n = \text{Law}(\overline{X}_n)$, as well as on the current mass process $\gamma_n(1)$. In contrast to the traditional McKean model, the dependency on the mass process induces a dependency of all the flow of measures η_p, for $0 \leq p \leq n$.

The mean field particle interpretation of this nonlinear measure-valued model is the Markov chain

$$(\gamma_n^N(1), \xi_n) \in (\mathbb{R}_+ \times E_n^N) \qquad \text{with} \quad \xi_n = \left(\xi_n^i\right)_{1 \leq i \leq N} \in E_n^N$$

with elementary transitions

$$\mathbb{P}\left(\xi_{n+1} \in dx \;\middle|\; (\gamma_n^N(1), \xi_n)\;\right) = \prod_{i=1}^{N} K_{n+1,\gamma_n^N}(\xi_n^i, dx^i) \tag{12.16}$$

$$\gamma_{n+1}^N(1) = \gamma_n^N(1)\; \eta_n^N(G_{n,\gamma_n^N}) \tag{12.17}$$

with the infinitesimal neighborhood $dx = dx^1 \times \ldots \times dx^N$ of a point $x = (x^1, \ldots, x^N) \in E_{n+1}^N$. In the above display, $\left(\gamma_n^N, \eta_n^N\right)$ stands for the pair of measures defined for any $f_n \in \mathcal{B}_b(E_n)$ by

$$\eta_n^N := \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_n^j} \quad \text{and} \quad \gamma_n^N(f_n) := \gamma_n^N(1) \times \eta_n^N(f_n)$$

The initial system ξ_0 consists of N independent and identically distributed random variables with common law η_0. We also assume that $\gamma_0^N(1) = \gamma_0(1) = \mu_0(1)$ is explicitly known.

12.3 Multiple-object nonlinear filtering equations

In this section we discuss in more detail the multiple-objects nonlinear filtering problems. These nonlinear evolution models in distribution spaces are particular examples of the measure valued evolution equations discussed in Section 12.2.1.

12.3.1 A partially observed branching model

A signal branching model

Suppose that at a given time n there are N_n^X targets $(X_n^i)_{1 \leq i \leq N_n^X}$, each taking values in some measurable state space E_n.

A target X_n^i, at time n, survives to the next time step with probability $s_n(X_n^i) \in [0,1]$, and it evolves to a new random state according to a given elementary Markov transition M'_{n+1}, from E_n into E_{n+1}.

In addition, any target X_n^i can spawn new targets at the next time, usually modeled by a spatial Poisson process with a given intensity measure $B_{n+1}(X_n^i, \cdot)$, on the state space E_{n+1}. At the same time, an independent collection of new targets is added to the scene. This additional and spontaneous branching process is often modeled by a spatial Poisson process with a prescribed intensity measure μ_{n+1} on E_{n+1}.

For any $n \geq 0$, and any $x_n \in E_n$, we set

$$\begin{aligned} \overline{\mu}_n(dx_n) &= \mu_n(dx_n)/\mu_n(1) \\ b_n(x_n) &= B_{n+1}(1)(x_n) \quad \text{and} \quad \overline{B}_{n+1}(x_n, dx_{n+1}) = \frac{B_{n+1}(x_n, dx_{n+1})}{B_{n+1}(1)(x_n)} \end{aligned}$$

The signal process can be interpreted as a branching process of the same form as the one discussed in Section 12.1.1. The occupation measures of the branching process are given by

$$\mathcal{X}_n := \sum_{1 \leq i \leq N_n^X} \delta_{X_n^i}$$

Using the same arguments as in the proof of Lemma 12.1.2, for any $f_n \in \mathcal{B}_b(E_n)$ we prove that

$$\mathbb{E}\left(\mathcal{X}_n(f_n) \mid \mathcal{X}_{n-1}\right) = \mathcal{X}_{n-1}Q_n(f_n) + \mu_n(f_n)$$

with the nonnegative integral operator

$$Q_n(f_n) = s_{n-1}\ M'_n(f_n) + b_{n-1}\ \overline{B}_n(f_n)$$

We notice that the operators Q_{n+1} can be rewritten in terms of the Feynman-Kac integral operator

$$Q_{n+1}(x_n, dx_{n+1}) = G_n(x_n)\ M_{n+1}(x_n, dx_{n+1}) \tag{12.18}$$

with the potential function

$$G_n(x_n) = s_n(x_n) + b_n(x_n)$$

and the Markov transition

$$M_{n+1}(x_n, dx_{n+1})$$

$$:= \frac{s_n(x_n)}{s_n(x_n) + b_n(x_n)}\ M'_{n+1}(x_n, dx_{n+1}) + \frac{b_n(x_n)}{s_n(x_n) + b_n(x_n)}\ \overline{B}_{n+1}(x_n, dx_{n+1})$$

This shows that we can use the mean field particle techniques developed in Section 12.1 to compute the intensity distribution flows associated with the signal branching process.

A partial observation model

Given a realization of the branching process $\mathcal{X}_n$ defined in Section 12.3.1, with a probability $d_n(x)$ every random target $X_n^i = x$ generates an observation Y_n^i, on some possibly different state space E_n^Y, with distribution $L_n(x, dy)$, where $L_n(x, dy)$ stands for some Markov transition from E_n into E_n^Y. Otherwise, with a probability $(1 - d_n(x))$, the target disappears from the scene, and goes into an auxiliary cemetery or coffin state c. The $[0, 1]$-valued function d_n is called the detection probability of the targets.

More formally, a given state x generates a random observation in the augmented state space $E_{n,c}^Y := E_n^Y \cup \{c\}$, with distribution

$$L_{n,c}(x, dy) := d_n(x)\ L_n(x, dy) + (1 - d_n(x))\ \delta_c(dy) \tag{12.19}$$

The resulting observation point process is the random measure

$$\mathcal{Y}_n = \sum_{1 \leq i \leq N_{n,c}^Y} \delta_{Y_n^i}$$

with $N_{n,c}^Y = N_n^X$, on the augmented state space $E_{n,c}^Y$.

In addition to this partial observation process, we also observe an additional, and independent of $(\mathcal{X}_p)_{p \leq n}$, Poisson point process

$$\mathcal{Y}'_n := \sum_{1 \leq i \leq N'_n} \delta_{Y_n'^i}$$

with intensity measure ν_n on E_n^Y.

We further assume that $\nu_n \ll \lambda_n$ and $L_n(x, .) \ll \lambda_n$, for any $x \in E_n$, for some reference measure $\lambda_n \in \mathcal{M}(E_n^Y)$. We also assume that the Radon-Nikodym derivatives given by

$$g_n(x, y) = \frac{dL_n(x, .)}{d\lambda_n}(y) \quad \text{and} \quad h_n(y) := \frac{d\nu_n}{d\lambda_n}(y) \tag{12.20}$$

are such that

$$h_n(y) + d_n(x) g_n(x, y) > 0$$

for any $(x, y) \in \left(E_n \times E_n^Y\right)$.

The full observation process on $E_{n,c}^Y$ is now given by the random measure

$$\mathcal{Y}_n'' = \mathcal{Y}_n + \mathcal{Y}_n'$$

The coffin state c being unobservable, the "real world" observation point process $\mathcal{Y}_n^o$ is the random measure on the state space E_n^Y, given by the trace $(\mathcal{Y}_n'')_{|E_n^Y}$ of the measure $\mathcal{Y}_n''$ on the set E_n^Y. More precisely, the observed random measure is given by the following formula

$$\mathcal{Y}_n^o := \left(\mathcal{Y}_n''\right)_{|E_n^Y} = (\mathcal{Y}_n)_{|E_n^Y} + \mathcal{Y}_n' \quad \text{with} \quad (\mathcal{Y}_n)_{|E_n^Y} := \sum_{1 \leq i \leq N_{n,c}^Y} 1_{E_n^Y}(Y_n^i)\ \delta_{Y_n^i} \tag{12.21}$$

Multi-target tracking problems are concerned with the sequential estimation of the random measures

$$\mathcal{X}_n = \sum_{1 \leq i \leq N_n^X} \delta_{X_n^i}$$

given the noisy and partial observation occupation measures

$$\mathcal{Y}_p^o = \sum_{1 \leq i \leq N_p^Y} \delta_{Y_p^i} \quad \text{with} \quad 0 \leq p \leq n$$

12.3.2 Probability hypothesis density equations

From the pure probabilistic viewpoint, multi-target tracking problems consist in estimating the conditional distributions of the occupation measures of spatial branching processes, given some noisy and partial observation random fields.

Besides the fact that the underlying signal is a well defined and an easy to sample Markov chain model $\mathcal{X}_n$, the computation of the likelihood of the observation process $\mathcal{Y}_n^o$, given the random state $\mathcal{X}_n$, involves intractable combinatorial calculations. In this section, we present a Poisson approximation model that simplifies drastically the analysis.

The equations associated with these approximated filters are expressed in terms of nonlinear evolution equations of intensity measures, of the same type as the one discussed in Section 12.2. For a single target filtering problem, these equations reduce to the traditional single target optimal filter equations.

We emphasize that the multiple-object filtering equations developed in this section are not optimal, in the sense that they only represent an "approximation" of the conditional distributions. The connections between these Poisson approximation models and the optimal filter are still an important, but difficult open research question.

Poisson approximation models

Further on in this section, we assume that the initial random measure $\mathcal{X}_0$ is a Poisson point process, with intensity measure $\gamma_0 = \mu_0 \in \mathcal{M}_+(E_0)$, on the initial state space E_0. We also consider the sequence of integral operators Q_n defined in (12.18).

Given a realization of $\mathcal{X}_0$, the corresponding observation process $\mathcal{Y}_0^o$ on E_0^Y is defined as in (12.21) with some detection functions d_0 on E_0, some clutter intensity measures ν_0, and some Markov transitions $(L_{c,0}, L_0)$ defined as in (12.19). We also assume that the regularity condition (12.20) is satisfied, for some reference measures λ_0 and some clutter intensity functions h_0.

For any function $f \in \mathcal{B}(E_0)$, we have the

$$\begin{aligned}\widehat{\gamma}_0(f) &:= \mathbb{E}\left(\mathcal{X}_0(f) \mid \mathcal{Y}_0^{\mathrm{o}}\right) \\ &= \gamma_0((1-\alpha_0)f) + \int \mathcal{Y}_0^{\mathrm{o}}(dy)\,(1-\beta_{0,\gamma_0}(y))\ \Psi_{\alpha_0 g_0(y,.)}(\gamma_0)(f)\end{aligned}$$

with the $[0,1]$-valued function β_γ on E_2 defined by

$$\beta_{0,\gamma_0}(y) := h_0(y)/\left[h_0(y) + \gamma_0(d_0 g_0(.,y))\right]$$

A detailed proof of this conditional formula relies on Poisson point process conditioning principles, and it is provided in Section 12.3.3 (cf. Corollary 12.3.12).

Suppose we have defined the measure valued process $(\widehat{\gamma}_p, \gamma_p)$ and the random signal-observation process $(\mathcal{X}_p, \mathcal{Y}_p^{\mathrm{o}})$, from the origin $p = 0$, up to a given time horizon $p = n$.

Given these values, we define the pair of random measures $(\mathcal{X}_{n+1}, \mathcal{Y}_{n+1}^{\mathrm{o}})$ as follows:

- Firstly, we let $\mathcal{X}_{n+1}$ be a spatial Poisson point process with intensity measure γ_{n+1} defined by the following recursions

$$\begin{aligned}\gamma_{n+1} &:= \widehat{\gamma}_n Q_n + \mu_n \\ \widehat{\gamma}_n(f) &:= \gamma_n((1-d_n)f) + \int \mathcal{Y}_n^{\mathrm{o}}(dy)\,(1-\beta_{\gamma_n}(y))\ \Psi_{d_n g_n(y,.)}(\gamma_n)(f)\end{aligned} \tag{12.22}$$

  for any function $f \in \mathcal{B}(E_n)$, with the $[0,1]$-valued parameters

$$\beta_{n,\gamma_n}(y) := \frac{h_n(y)}{[h_n(y) + \gamma_n(d_n g_n(.,y))]}$$

- Given a realization of $\mathcal{X}_{n+1}$, the corresponding observation process $\mathcal{Y}_{n+1}^{\mathrm{o}}$ is defined as in (12.21), for some detection $[0,1]$-valued functions d_{n+1} on E_{n+1}, some clutter intensity measures ν_{n+1}, and some Markov transitions $(L_{c,(n+1)}, L_{n+1})$ defined as in (12.19); and satisfying (12.20), for some reference measures λ_{n+1} and some functions h_{n+1}.

We let $\mathcal{F}_n^Y = \sigma\left(\mathcal{Y}_p^{\mathrm{o}},\ 0 \leq p \leq n\right)$ be the filtration generated by the observation point processes $\mathcal{Y}_p^{\mathrm{o}}$, from the origin $p = 0$, up to the current time $p = n$. By construction, for any function $f \in \mathcal{B}(E_{n+1})$, we clearly have that

$$\mathbb{E}\left(\mathcal{X}_{n+1}(f) \mid \mathcal{F}_n^Y\right) = \gamma_{n+1}(f)$$

In addition, using the same arguments as the ones we used at the initial time $n = 0$, we have the updating formulae

$$\begin{aligned}&\widehat{\gamma}_{n+1}(f) \\ &:= \mathbb{E}\left(\mathcal{X}_{n+1}(f) \mid \mathcal{F}_{n+1}^Y\right) \\ &= \gamma_{n+1}((1-d_{n+1})f) + \int \mathcal{Y}_{n+1}^{\mathrm{o}}(dy)\left(1-\beta_{\gamma_{n+1}}(y)\right)\ \Psi_{d_{n+1}g_{n+1}(y,.)}(\gamma_{n+1})(f)\end{aligned}$$

In summary, we have proved that the solution of the (PHD) Equations (12.22) coincides at any time step, with the desired conditional distributions

$$\widehat{\gamma}_n(f) = \mathbb{E}\left(\mathcal{X}_n(f) \mid \mathcal{F}_n^Y\right) \quad \text{and} \quad \mathbb{E}\left(\mathcal{X}_n(f) \mid \mathcal{F}_{n-1}^Y\right) = \gamma_n(f)$$

We end this section with a more synthetic description of the PHD equations. More precisely, using the decomposition (12.18), we can rewrite the PHD Equation (12.22) in terms of a nonlinear model of the form (12.11), combining in a single step the updating $\gamma_n \rightsquigarrow \widehat{\gamma}_n$ and the prediction transition $\widehat{\gamma}_n \rightsquigarrow \gamma_{n+1}$.

Proposition 12.3.1 *The PHD filter satisfies the integral Equation (12.11), with the integral operator given by*

$$Q_{n+1,\gamma_n}(x_n, dx_{n+1}) = g_{n,\gamma_n}(x_n) M_{n+1}(x_n, dx_{n+1}) + \gamma_n(1)^{-1}\ \mu_{n+1}(dx_{n+1}) \tag{12.23}$$

The likelihood function g_{n,γ_n} is given by

$$g_{n,\gamma_n} := r_n \times \widehat{g}_{n,\gamma_n} \quad \textit{with} \quad r_n := (s_n + b_n) \tag{12.24}$$

and

$$\widehat{g}_{n,\gamma_n}(x_n) := (1 - d_n(x_n)) + d_n(x_n) \int \mathcal{Y}_n^{\mathrm{o}}(dy)\ \frac{g_n(x_n, y_n)}{h_n(y_n) + \gamma_n(d_n g_n(\boldsymbol{.}, y_n))}$$

An updating-prediction formulation

In this section, we present a more traditional updating-prediction formulation of the PHD equations presented in (12.22). These updating-prediction models are often used in the literature of multiple target tracking.

We extend the observation state space E_n^Y by adding a virtual but cemetery type state $\{c\}$, and we consider the following likelihood functions on $E_{n,c}^Y = E_n^Y \cup \{c\}$

$$\widehat{g}_{n,\gamma}^c(., y) = \begin{cases} (1 - d_n) & \text{if} \quad y = c \\ \dfrac{d_n g_n(., y_n)}{h_n(y) + \gamma(d_n g_n(., y))} & \text{if} \quad y \neq c \end{cases}$$

In this interpretation, the state $y = c$ is considered as a virtual observable state, with a likelihood function $\widehat{g}_{n,\gamma}^c(x, c) = (1 - d_n(x))$ that measures the undetectability properties of the site x. The likelihood function is high in regions with low detectability conditions.

In this notation, we have

$$\gamma_{n+1} = \widehat{\gamma}_n Q_{n+1} + \mu_{n+1} \quad \text{with} \quad Q_{n+1}(f) := r_n\ M_{n+1}(f)$$

with the updated measures defined below

$$\widehat{\gamma}_n(f) := \gamma_n(\widetilde{g}_{n,\gamma_n}^c f) \quad \text{with} \quad \widetilde{g}_{n,\gamma_n}^c = \int \mathcal{Y}_n^c(dy)\ \widehat{g}_{n,\gamma_n}^c(., y)$$

and $\mathcal{Y}_n^c = \mathcal{Y}_n^{\mathrm{o}} + \delta_c$. Notice that

$$\widehat{\gamma}_n(1) = \gamma_n(\widetilde{g}_{n,\gamma_n}^c) \quad \text{and} \quad \widehat{\eta}_n(dx) := \widehat{\gamma}_n(dx)/\widehat{\gamma}_n(1) = \Psi_{\widetilde{g}_{\gamma_n,n}^c}(\eta_n)(dx)$$

from which we find the recursive formulae

$$\begin{pmatrix} \gamma_n(1) \\ \eta_n \end{pmatrix} \xrightarrow{\text{updating}} \begin{pmatrix} \widehat{\gamma}_n(1) \\ \widehat{\eta}_n \end{pmatrix} \xrightarrow{\text{prediction}} \begin{pmatrix} \gamma_{n+1}(1) \\ \eta_{n+1} \end{pmatrix}$$

The updating transition is defined by

$$\widehat{\gamma}_n(1) = \gamma_n(1)\ \eta_n(\widetilde{g}_{n,\gamma_n}^c) \quad \text{and} \quad \widehat{\eta}_n = \Psi_{\widetilde{g}_{\gamma_n,n}^c}(\eta_n)$$

It is instructive to observe that for fully detectable target models without clutter, we have

$$(d_n, h_n) = (1, 0) \Rightarrow \widehat{g}_{n,\gamma_n}^c(x, y) = 1_{E_n^Y}(y)\ \frac{g_n(x, y_n)}{\gamma_n(g_n(., y))} \Rightarrow \widehat{\gamma}_n(1) = \gamma_n(1)$$

The prediction transitions are given by

$$\gamma_{n+1}(1) = \widehat{\gamma}_n(r_n) + \mu_{n+1}(1) \quad \text{and} \quad \eta_{n+1} = \Psi_{r_n}(\widehat{\eta}_n)\, M'_{n+1,\widehat{\gamma}_n}$$

In the above displayed formula, $M'_{n+1,\gamma}$ is the collection of Markov transitions defined by

$$M'_{n+1,\gamma}(x,.) = \alpha'_n(\gamma)\ M_{n+1}(x,.) + \left(1-\alpha'_n(\gamma)\right)\ \overline{\mu}_{n+1}$$

with the collection of $[0,1]$-valued parameters

$$\alpha'_n(\gamma) = \gamma(r_n)/(\gamma(r_n)+\mu_{n+1}(1))$$

It is important to mention that the updating, as well as the prediction transitions, can be rewritten in terms of a measure valued equation of the same form as the one presented in Section 12.2.1. For instance, we can decompose the two step updating-prediction discussed above with intermediate time steps between integers

$$\gamma_n \quad \longrightarrow \quad \gamma_{n+1/2} := \widehat{\gamma}_n \quad \longrightarrow \quad \gamma_{n+1}$$

In this notation, we have a couple of one step transformations

$$\gamma_{n+1/2} = \Xi_{n+1/2}(\gamma_n) \quad \text{and} \quad \gamma_{n+1} = \Xi_{n+1}(\gamma_{n+1/2})$$

with the one step transformations, defined for any $\gamma \in \mathcal{M}_+(E_n)$ by

$$\begin{array}{rcl} \Xi_{n+1/2}(\gamma) & = & \gamma(\widetilde{g}^c_{n,\gamma})\ \Psi_{\widetilde{g}^c_{n,\gamma}}(\gamma) \\ \Xi_{n+1}(\gamma) & = & [\gamma(r_n)+\mu_{n+1}(1)]\ \Psi_{r_n}(\gamma)\, M'_{n+1,\gamma} \end{array}$$

Now, it should be clear that the updating and the prediction transitions can be approximated using mean field particle models presented in Section 12.2.2. Therefore, the performances of these updating-prediction mean field algorithms are direct consequences of the convergence analysis of the general particle model discussed in Section 12.2.2.

12.3.3 Spatial Poisson processes

In this short section, we recall some more or less well known results on spatial Poisson point processes, including restriction techniques and conditioning principles for partially observed models.

Preliminary results

We consider a measure $\gamma \in \mathcal{M}(E)$ on some measurable state space E s.t. $\gamma(1) > 0$. We let N be a integer valued Poisson random variable with parameter $\gamma(1)$. We also denote by $X = (X^i)_{i\geq 1}$ a sequence of independent and identically distributed random variables with common distribution $\eta(dx) := \gamma(dx)/\gamma(1)$. We assume that N and X are independent.

Definition 12.3.2 *The Poisson point process $\mathcal{X}$ with intensity measure γ is the random measure defined below*

$$\mathcal{X} := m_N(X) = \sum_{1\leq i\leq N} \delta_{X^i} \in \mathcal{P}(E)$$

One of the main simplifications of Poisson point processes comes from the fact that their expectation measure coincides with their intensity measures:

$$\mathbb{E}\left(\mathcal{X}(f)\right) = \mathbb{E}\left(\mathbb{E}\left(\mathcal{X}(f)\mid N\right)\right) = \mathbb{E}\left(N\eta(f)\right) = \gamma(1)\eta(f) = \gamma(f)$$

Definition 12.3.3 *For every sequence of point $x = (x^i)_{i\geq 1}$ in E, any $A \in \mathcal{E}$, and every $p \geq 0$, we denote by $m_{p,A}(x)$ the restriction of the occupation measure $m_p(x)$ to the set A.*

$$m_{p,A}(x)(dy) = m_p(x)(dy)1_A(y) = \sum_{1\leq i\leq p} 1_A(x^i)\delta_{x^i}(dy)$$

.

Notice that

$$m_p(x)(A) = \sum_{1\leq i\leq p} 1_{A_i}(x^i) > 0 \implies m_{p,A}(x) = m_p(x)(A)\ \Psi_{1_A}(m(x))$$

with the Boltzmann-Gibbs transformation Ψ_G associated with the indicator function $G = 1_A$, defined in (4.31).

Lemma 12.3.4 *Let $(\mathcal{X}_j)_{j\geq 1}$ be a sequence of independent Poisson point processes with intensity measure $(\gamma_i)_{i\geq 1}$ on some common measurable state space E. For any $d \geq 1$, $\mathcal{X}$ is a Poisson point process with intensity measure $\sum_{1\leq i\leq d}\gamma_i$ if, and only if, $\mathcal{X}$ is equal in law to the Poisson point process $\sum_{1\leq i\leq d}\mathcal{X}_i$.*

Proof:
By symmetry arguments, we have for any $F \in \mathcal{B}(\mathcal{M}(E))$ and any $d \geq 1$

$$\mathbb{E}\left(F\left(\sum_{1\leq i\leq d}\mathcal{X}_i\right)\right) =$$

$$e^{-\sum_{1\leq i\leq d}\gamma_i(1)}\sum_{p_1,\ldots,p_d\geq 0}\frac{1}{p_1!\ldots p_d!}\int F\left(\sum_{1\leq i\leq d} m_{p_i}(x_i)\right)\ \prod_{1\leq i\leq d}\gamma_i^{\otimes p_i}(dx_i)$$

This implies that

$$\begin{aligned}&\mathbb{E}\left(F\left(\sum_{1\leq i\leq d}\mathcal{X}_i\right)\right) =\\ &= e^{-\sum_{1\leq i\leq d}\gamma_i(1)}\sum_{s\geq 0}\frac{1}{s!}\\ &\qquad \sum_{p_1+\ldots+p_d=s}\frac{s!}{p_1!\ldots p_d!}\int F\left(m_s(x)\right)\ \left(\gamma_1^{\otimes p_1}\otimes\ldots\otimes\gamma_d^{\otimes p_d}\right)(dx)\end{aligned}$$

In the above displayed integral $dx = dx^1\times\ldots\times dx^s$ stands for an infinitesimal neighborhood of the point $x = (x^i)_{1\leq i\leq s}$. This implies that

$$\mathbb{E}\left(F\left(\sum_{1\leq i\leq d}\mathcal{X}_i\right)\right) = e^{-\sum_{1\leq i\leq d}\gamma_i(1)}\sum_{s\geq 0}\frac{1}{s!}\int F\left(m_s(x)\right)\ \left(\sum_{i=1}^{d}\gamma_i\right)^{\otimes s}(dx) \tag{12.25}$$

This shows that $\sum_{1\leq i\leq d}\mathcal{X}_i$ is a Poisson point process with intensity measure $\sum_{1\leq i\leq d}\gamma_i$. In addition, by (12.25), any Poisson point process with such an intensity measure has the same law as $\sum_{1\leq i\leq d}\mathcal{X}_i$. ■

The next result is a direct consequence of Lemma 12.3.4.

Lemma 12.3.5 *Let $\mathcal{X} := \sum_{1\leq i\leq N}\delta_{X^i}$ be a Poisson point process with intensity measure γ that is the random measure on E. We consider a measurable subset $A \subset E$, such that $\gamma(A) > 0$. Then, the restriction, or the trace, $\mathcal{X}_A = m_{N,A}(X)$ of $\mathcal{X}$ on the set A is again a Poisson point process with intensity measure $\gamma_A(dx) := 1_A(x)\gamma(dx)$.*

In addition, the conditional distribution of $\mathcal{X}$ given $\mathcal{X}_A$ is given for any $F \in \mathcal{B}(\mathcal{M}(E))$ by the following formula

$$\mathbb{E}\left(F(\mathcal{X})\,|\mathcal{X}_A\right) = e^{-\gamma(A^c)}\sum_{p\geq 0}\ \frac{1}{p!}\ \int F\left(\mathcal{X}_A + m_p(x)\right)\ \gamma_{A^c}^{\otimes p}(dx)$$

In the above, displayed integral $dx = dx^1\times\ldots\times dx^p$ stands for an infinitesimal neighborhood of the point $x = (x^i)_{1\leq i\leq p}$.

Proof:
Using the decomposition

$$\gamma(dx) = 1_A(x)\gamma(dx) + 1_{A^c}(x)\gamma(dx) \;\Rightarrow\; \gamma = \gamma_A + \gamma_{A^c}$$

for any $F \in \mathcal{B}(\mathcal{M}(E))$ we find that

$$\mathbb{E}\left(F\left(\mathcal{X}_A\right)\right) = e^{-\gamma(1)} \sum_{s\geq 0} \frac{1}{s!} \int F\left(m_{s,A}(x)\right) \; \left(\gamma_A + \gamma_{A^c}\right)^{\otimes s}(dx)$$

By symmetry arguments, this implies that

$$\mathbb{E}\left(F\left(\mathcal{X}_A\right)\right) = e^{-\gamma(1)} \sum_{s\geq 0} \frac{1}{s!} \sum_{p+q=s} \frac{s!}{p!q!} \int F\left(m_{s,A}(x)\right) \; \left[\gamma_A^{\otimes p} \otimes \gamma_{A^c}^{\otimes (s-p)}\right](dx)$$

from which we find that

$$\begin{aligned}\mathbb{E}\left(F\left(\mathcal{X}_A\right)\right) &= e^{-\gamma(1)} \sum_{p\geq 0} \frac{1}{p!} \left(\sum_{s\geq p} \frac{\gamma(A^c)^{s-p}}{(s-p)!}\right) \int F\left(m_p(x)\right) \gamma_A^{\otimes p}(dx) \\ &= e^{-(\gamma(E)-\gamma(A^c))} \sum_{p\geq 0} \frac{1}{p!} \int F\left(m_p(x)\right) \gamma_A^{\otimes p}(dx)\end{aligned}$$

The last assertion is a direct consequence of Lemma 12.3.4, applied to $d = 2$, replacing $(\mathcal{X}_1, \mathcal{X}_2)$ by $(\mathcal{X}_A, \mathcal{X}_{A^c})$. This ends the proof of the lemma. ∎

Some conditioning principles

We consider a measure $\gamma \in \mathcal{M}(E_1)$ on some measurable space $(E_1, \mathcal{E}_1)$ and a bounded positive integral operator Q from $(E_1, \mathcal{E}_1)$ into an auxiliary measurable space $(E_2, \mathcal{E}_2)$.

We further assume that $Q(1) > 0$, γ-a.e., and we let

$$\mathcal{X} := m_N(X_1, X_2) = \sum_{1\leq i\leq N} \delta_{(X_1^i, X_2^i)} \tag{12.26}$$

be the Poisson point process on some product space $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ with intensity measure Γ of the following form

$$\Gamma(d(x_1, x_2)) := \gamma(dx_1)\; Q(x_1, dx_2)$$

It is immediate to check that the marginal random measures $\mathcal{X}_j := m_N(X_j)$ are Poisson point processes on E_j, $j = 1, 2$, with intensity measures

$$\gamma_1(dx) := Q(1)(x)\; \gamma(dx) \quad \text{and} \quad \gamma_2 := \gamma Q$$

Our next objective is to describe the conditional distributions of the random measures $\mathcal{X}_i$ w.r.t. $\mathcal{X}_j$, with $i \neq j$.

Lemma 12.3.6 *For any $f \in \mathcal{B}(E_1)$ we set $\gamma_f(dx) := f(x)\; \gamma(dx)$. The integral operators*

$$\overline{Q}(x_1, dx_2) = \frac{Q(x_1, dx_2)}{Q(x_1, E_2)} \quad \textit{and} \quad f \in \mathcal{B}(E_1) \mapsto Q_\gamma(f) := \frac{d\gamma_f Q}{d\gamma Q}$$

are well defined Markov transitions from E_1 into E_2, resp. from E_2 into E_1. In addition, if we have $Q(x_1, .) \ll \gamma Q$, for every $x_1 \in E_1$, then we have the following explicit formula

$$Q_\gamma(x_2, dx_1) := \gamma(dx_1)\; \frac{dQ(x_1, .)}{d\gamma Q}(x_2)$$

Proof:

The fact that $\overline{Q}$ is a well defined Markov transition is immediate. To check the second assertion, we use the fact that

$$\gamma_f(dx) := f(x)\ \gamma(dx) \ll \gamma(dx) \Longrightarrow \gamma_f Q \ll \gamma Q$$

for any $f \in \mathcal{B}(E_1)$. The r.h.s. assertion comes from the following series of implications

$$\begin{aligned} \gamma Q(A) = 0 \quad &\Rightarrow \quad Q(1_A) = 0 \quad \gamma - \text{almost everywhere} \\ &\Rightarrow \quad Q(1_A) = 0 \quad \gamma_f - \text{almost everywhere} \Rightarrow \gamma_f Q(A) = 0 \end{aligned}$$

Using this property, we define the following operator from $\mathcal{B}(E_1)$ into $\mathcal{B}(E_2)$:

$$\forall f \in \mathcal{B}(E_1) \quad \forall x \in E_2 \qquad Q_\gamma(f)(x_2) := \frac{d\gamma_f Q}{d\gamma Q}(x_2)$$

Notice that $Q_\gamma(1)(x_2) = 1$, and for any $(f, g) \in \mathcal{B}(E_1)^2$, we have

$$\gamma_{f+g} = \gamma_f + \gamma_g \Rightarrow Q_\gamma(f+g) = Q_\gamma(f) + Q_\gamma(g)$$

Using the fact that $\lim_{n\to\infty} \gamma_{1_{A_n}} = 0$ for every decreasing sequence of subsets $A_n \in E_1$ s.t. $\lim_{n\to\infty} A_n = \emptyset$, we prove that $\lim_{n\to\infty} Q_\gamma(1_{A_n})(x_2) = 0$, γQ-a.e. This implies that $A \in \mathcal{E}_1 \mapsto Q_\gamma(1_A)(x_2)$ is a well defined probability measure $Q_\gamma(x_2, dx_1)$ on the set $(E_1, \mathcal{E}_1)$, and we have the following γQ-a.e. Lebesgue integral representation

$$Q_\gamma(f)(x_2) = \int Q_\gamma(x_2, dx_1) f(x_1)$$

This ends the proof of the lemma. ∎

By construction, we have the equivalent time reversal formulae

$$\gamma(dx_1)\ Q(x_1, dx_2) = (\gamma Q)\,(dx_2)\ Q_\gamma(x_2, dx_1) \tag{12.27}$$

Notice that

$$\gamma(dx_1)\ Q(x_1, dx_2) = \gamma(G)\ \Psi_G(\eta)(dx_1)\ \overline{Q}(x_1, dx_2)$$

with

$$\eta(dx_1) = \gamma(dx_1)/\gamma(1) \quad \text{and} \quad G := Q(1)$$

In the same vein, we also have that

$$(\gamma Q)\,(dx_2)\ Q_\gamma(x_2, dx_1) = \gamma(G)\ (\Psi_G(\eta)\overline{Q})(dx_2)\ Q_\gamma(x_2, dx_1)$$

This implies that

$$\Psi_G(\eta)(dx_1)\ \overline{Q}(x_1, dx_2) = (\Psi_G(\eta)\overline{Q})(dx_2)\ Q_\gamma(x_2, dx_1)$$

from which we conclude that $Q_\gamma = \overline{Q}_{\Psi_G(\eta)}$, where $\overline{Q}_\gamma$ is defined as Q_γ by replacing Q by $\overline{Q}$.

Using rather elementary manipulations, we prove the following lemma.

Lemma 12.3.7 *For any functions $F_j \in \mathcal{B}(\mathcal{M}(E_j))$, with $j = 1, 2$, we have the almost sure formulae:*

$$\mathbb{E}\,(F_1(\mathcal{X}_1)\ \mid \mathcal{X}_2\,) = \int F_1\,(m_N(x_1))\ \prod_{1\le i\le N} Q_\gamma(X_2^i, dx_1^i)$$

and

$$\mathbb{E}\,(F_2(\mathcal{X}_2)\ \mid \mathcal{X}_1\,) = \int F_2\,(m_N(x_2))\ \prod_{1\le i\le N} \overline{Q}(X_1^i, dx_2^i)$$

In the above displayed formula $\overline{Q}$ and Q_γ stand for the pair of Markov transitions introduced in Lemma 12.3.6.

Proof:
To prove the second assertion, we recall that $\mathcal{X}_1 := m_N(X_1) = \sum_{1\leq i\leq N} \delta_{X_1^i}$ is a Poisson point process on E_1, with intensity measure $\gamma_1(dx) := Q(1)(x)\ \gamma(dx)$. From this result, we find that

$$\mathbb{E}\left(F_1(\mathcal{X}_1)\ \left\{\int F_2\left(m_N(x_2)\right)\ \prod_{1\leq i\leq N} \overline{Q}(X_1^i, dx_2^i)\right\}\right)$$

$$= e^{-\gamma Q(1)} \sum_{p\geq 0} \tfrac{1}{p!}$$

$$\times \int F_1\left(m_p(x_1)\right)\ F_2\left(m_p(x_2)\right) \prod_{1\leq i\leq p} \left[\overline{Q}(x_1^i, dx_2^i)\ Q(1)(x_1^i)\ \gamma(dx_1^i)\right]$$

$$= e^{-\gamma Q(1)} \sum_{p\geq 0} \tfrac{1}{p!} \int F_1\left(m_p(x_1)\right)\ F_2\left(m_p(x_2)\right) \prod_{1\leq i\leq p} \left[\gamma(dx_1^i) Q(x_1^i, dx_2^i)\right]$$

$$= \mathbb{E}\left(F_1(\mathcal{X}_1) F_2(\mathcal{X}_2)\right)$$

This ends the proof of the second assertion. Using the time reversal decomposition formula (12.27), and recalling that Q_γ is a Markov transition, the first assertion is a direct consequence of the second one. This ends the proof of the lemma. ∎

Partially observed models

We consider a spatial branching signal model defined by a Poisson point process $\mathcal{X} := \sum_{1\leq i\leq N} \delta_{X^i}$, with intensity measure γ on some measurable state space $(E_1, \mathcal{E}_1)$, and we set $\eta(f) := \gamma(f)/\gamma(1)$, for any $f \in \mathcal{B}(E_1)$.

The random variable $\mathcal{X}$ is partially observed, on some possibly different measurable state space E_2. The observation is defined by a spatial point process. It consists in a collection of random observation variables, directly generated by some random points in the support of $\mathcal{X}$, plus some random observations unrelated to $\mathcal{X}$, sometimes called the clutter. We use the partial observation model presented in Section 12.3.1.

For the convenience of the reader, we briefly recall the description of this model in this static framework. Given a realization of $\mathcal{X}$, every random state $X^i = x$ generates an observation Y^i on $E_2 \cup \{c\}$ with distribution

$$L_c(x, dy) := d(x)\ L(x, dy) + (1 - d(x))\ \delta_c(dy)$$

The function d represents the detectability degree of the states, and $L(x, .)$ stands for the distribution of the random observations on E_2 generated by the point x in E_1. The resulting observation process is the random measure $\mathcal{Y} = \sum_{1\leq i\leq N} \delta_{Y^i}$ on the augmented state space $E_2 \cup \{c\}$.

In addition to this partial observation model, we also observe an additional, and independent of $\mathcal{X}$, clutter Poisson point process $\mathcal{Y}' := \sum_{1\leq i\leq N'} \delta_{Y'^i}$, with intensity measure ν on E_2. As in (12.19), we further assume that ν and $L(x, .) \ll \lambda$, for some $\lambda \in \mathcal{M}(E_2)$, and we set

$$g(x, y) = \frac{dL(x, .)}{d\lambda}(y) \quad \text{and} \quad h := d\nu/d\lambda \tag{12.28}$$

We also suppose that $h(y) + \gamma(dg(., y)) > 0$, for any $y \in E_2$.

The full observation process on $E_2 \cup \{c\}$ is given by the random measure $\mathcal{Y}'' = \mathcal{Y} + \mathcal{Y}'$, while the "real world" observation $\mathcal{Y}^{\mathrm{o}}$ is given the random measure

$$\mathcal{Y}^{\mathrm{o}} := \mathcal{Y}''_{|E_2} = \mathcal{Y}_{|E_2} + \mathcal{Y}' \Longrightarrow \mathcal{Y} + \mathcal{Y}' = \mathcal{Y}^{\mathrm{o}} + N_c\ \delta_c \quad \text{with} \quad N_c := \mathcal{Y}(\{c\}) \tag{12.29}$$

The following proposition results from the construction of the observation process $\mathcal{Y}''$ on $E_2 \cup \{c\}$.

Proposition 12.3.8 *A version of the conditional distribution of the random measure $\mathcal{Y}''$ given $\mathcal{X} := \sum_{1\leq i\leq N} \delta_{X^i}$ is given for any function $F \in \mathcal{B}(\mathcal{M}(E_2 \cup \{c\}))$ by*

$$\mathbb{E}\left(F\left(\mathcal{Y}''\right)|\mathcal{X}\right)$$

$$= e^{-\nu(1)} \sum_{p\geq 0} \tfrac{1}{p!} \int F\left(m_p(y') + m_N(y)\right)\ \nu^{\otimes p}(dy') \quad \prod_{1\leq i\leq N} L_c(X^i, dy^i)$$

Our next objective is to compute the conditional distribution of $\mathcal{X}$, given the observation process $\mathcal{Y}^{\mathrm{o}}$.

Definition 12.3.9 *We let $\mathcal{Z}$ be the spatial point process defined by*

$$\mathcal{Z} := \sum_{1\leq i\leq N} \delta_{(X^i, Y^i)} + \sum_{1\leq i\leq N'} \delta_{(c, Y'^i)} := \sum_{1\leq i\leq N''} \delta_{(Z_1^i, Z_2^i)} \tag{12.30}$$

For any $i \in \{1, 2\}$, we denote by $\mathcal{B}_b(E_i \cup \{c\})$ the set of functions $f \in \mathcal{B}_b(E_i)$, extended to $E_i \cup \{c\}$, by setting $f(c) = 0$.

We observe that

$$\begin{aligned}
\mathcal{Z}_1 &= \sum_{1\leq i\leq N''} \delta_{Z_1^i} = \mathcal{X} + N'\delta_c \\
\mathcal{Z}_2 &= \sum_{1\leq i\leq N''} \delta_{Z_2^i} = (\mathcal{Z}_2)_{|E_2} + (\mathcal{Z}_2)_{|\{c\}} = \mathcal{Y}^o + N_c\delta_c
\end{aligned}$$

By construction, the random measure $\mathcal{Z}$ is a Poisson point process taking values in the state space

$$E_c = [(E_1 \cup \{c\}) \times (E_2 \cup \{c\})]$$

with intensity distribution given by the factorization formulae

$$\gamma(dx)L_c(x, dy) + \delta_c(dx)\nu(dy)$$

$$= \left[\underbrace{\gamma(dx)\ 1_{E_1}(x) + \nu(1)\ \delta_c(dx)}_{=\gamma_c(dx)}\right] \left[\underbrace{1_{E_1}(x)\ L_c(x, dy) + 1_c(x)\frac{\nu(dy)}{\nu(1)}}_{M_c(x,dy)}\right]$$

The marginal of the above distribution w.r.t. the second component is given by

$$\begin{aligned}
\gamma_c M_c(dy) &= (\gamma L_c + \nu)(dy) \\
&= [\gamma(dg(.,y)) + h(y)]\ \lambda(dy) + \gamma(1-d)\ \delta_c(dy)
\end{aligned}$$

On the other hand we have the decomposition

$$\gamma_c(dx)\ M_c(x, dy)$$

$$= [\gamma(dx)\ d(x)\ g(x, y) + \delta_c(dx)h(y)]\ \lambda(dy) + \gamma(dx)(1 - d(x))\ \delta_c(dy)$$

This yields the Bayes' type formula

$$\gamma_c(dx)\ M_c(x,dy) = \gamma_c M_c(dy)\ M_{c,\gamma_c}(y,dx)$$

with the Markov transition

$$M_{c,\gamma_c}(y,dx)$$

$$:= 1_{E_2}(y)\ \frac{[\gamma(dx)\ d(x)\ g(x,y) + \delta_c(dx)h(y)]}{[\gamma(dg(.,y)) + h(y)]} \quad + \quad 1_c(dy)\ \frac{\gamma(dx)(1-d(x))}{\gamma(1-d)}$$

This implies that

$$M_{c,\gamma_c}(y,dx) = 1_{E_2}(y)\ Q_{\gamma_c}(y,dx) + 1_c(dy)\ \Psi_{1-d}(\eta)(dx) \tag{12.31}$$

with the $[0,1]$-valued function β_γ on E_2 defined by

$$\beta_\gamma(y) := h(y)/\left[h(y) + \gamma(dg(.,y))\right] \tag{12.32}$$

and the Markov transition Q_{γ_c} from E_2 into $E_1 \cup \{c\}$ defined by the following formula

$$Q_{\gamma_c}(y,dx) = (1-\beta_\gamma(y))\ \Psi_{dg(y,.)}(\eta)(dx) + \beta_\gamma(y)\ \delta_c(dx) \tag{12.33}$$

We summarize the above discussion with the following proposition.

Proposition 12.3.10 *A version of the conditional distribution of the random measure $\mathcal{Z}_1$ given $\mathcal{Z}_2$ is given for any function $F \in \mathcal{B}(\mathcal{M}(E_1 \cup \{c\}))$ by*

$$\mathbb{E}\left(F(\mathcal{Z}_1) \mid \mathcal{Z}_2\right) = \int F\left(m_{\mathcal{Z}_2(1)}(x)\right) \prod_{1 \le i \le \mathcal{Z}_2(1)} M_{c,\gamma_c}(\mathcal{Z}_2^i, dx^i)$$

with the Markov M_{c,γ_c} transition, from $E_2 \cup \{c\}$ into $E_1 \cup \{c\}$, defined in (12.31).

Using the fact that

$$\mathcal{Z}_2 = (\mathcal{Z}_2)_{|E_2} + (\mathcal{Z}_2)_{|\{c\}} = \mathcal{Y}^o + N_c\delta_c$$

for any function $F \in \mathcal{B}(\mathcal{M}(E_2 \cup \{c\}))$, we also prove the following equation

$$\mathbb{E}\left(F(\mathcal{Z}_2) \mid \mathcal{Y}^{\mathrm{o}}\right) = e^{-\gamma(1-d)} \sum_{p \ge 0} \frac{\gamma(1-d)^p}{p!}\ F\left(p\delta_c + \mathcal{Y}^{\mathrm{o}}\right)$$

If we set

$$\mathcal{Y}^{\mathrm{o}} := \sum_{1 \le i \le N_1} \delta_{Y_1^i} = \sum_{1 \le i \le N} 1_{E_2}(Y^i)\ \delta_{Y^i} + \mathcal{Y}'$$

then we find that

$$\mathbb{E}\left(F(\mathcal{Z}_1) \mid (\mathcal{Z}_2)_{|E_2}\right)$$

$$= \mathbb{E}\left(\mathbb{E}\left(F(\mathcal{Z}_1) \mid \mathcal{Z}_2\right) \mid (\mathcal{Z}_2)_{|E_2}\right)$$

$$= e^{-\gamma(1-d)} \sum_{p \ge 0} \frac{\gamma(1-d)^p}{p!}$$

$$\times \int F\left(m_p(x') + m_{\mathcal{Y}^o(1)}(x)\right)\ \Psi_{(1-d)}(\eta)^{\otimes p}\left(dx'\right) \prod_{i=1}^{\mathcal{Y}^o(1)} Q_{\gamma_c}\left(Y_1^i, dx^i\right)$$

We summarize the above discussion with the following theorem.

Theorem 12.3.11 *A version of the conditional distribution of*

$$\mathcal{X}_c := \mathcal{X} + N'\delta_c$$

w.r.t. the observation process $\mathcal{Y}^{\mathrm{o}} = \sum_{1\leq i\leq N_1}\delta_{Y_1^i}$ *is given for any function* $F \in \mathcal{B}\left(\mathcal{M}(E_1\cup\{c\})\right)$ *by*

$$\mathbb{E}\left(F\left(\mathcal{X}_c\right)|\mathcal{Y}^{\mathrm{o}}\right)$$

$$= e^{-\gamma(1-d)}\sum_{p\geq 0}\frac{\gamma(1-d)^p}{p!}\int F\left(m_p(x')+m_{N_1}(x)\right)\ \Psi_{(1-d)}(\eta)^{\otimes p}\left(dx'\right)\ \prod_{i=1}^{N_1}Q_{\gamma_c}\left(Y_1^i,dx^i\right)$$

with the Markov Q_{γ_c} *transition, from* E_2 *into* $E_1\cup\{c\}$*, defined in (12.33).*

The conditional expectation measures of the random point processes $\mathcal{Y}_c$ and $\mathcal{Y}^{\mathrm{o}}$ resp. $\mathcal{X}_c$ and $\mathcal{X}$, given the point process $\mathcal{X}$, resp. $\mathcal{Y}^{\mathrm{o}}$, are now easily computed.

Corollary 12.3.12 *For any function* $f\in\mathcal{B}(E_2\cup\{c\})$*, we have the almost sure integral representation formula*

$$\mathbb{E}\left(\mathcal{Y}_c(f)\mid\mathcal{X}\right) \;=\; \mathbb{E}\left(\mathcal{Y}^{\mathrm{o}}(f)\mid\mathcal{X}\right) = \mathcal{X}(dL(f))+\nu(f)$$

and for any function $f\in\mathcal{B}(E_1\cup\{c\})$ *we have the almost sure integral representation formula*

$$\begin{aligned}\mathbb{E}\left(\mathcal{X}_c(f)\mid\mathcal{Y}^{\mathrm{o}}\right) &= \mathbb{E}\left(\mathcal{X}(f)\mid\mathcal{Y}^{\mathrm{o}}\right)\\ &= \gamma((1-d)f)+\int\mathcal{Y}^{\mathrm{o}}(dy)\left(1-\beta_\gamma(y)\right)\ \Psi_{dg(y,\bullet)}(\eta)(f)\end{aligned}$$

with the $[0,1]$*-valued function* β_γ *defined in (12.32).*

In particular, the conditional mean value of the number of states N *given the spatial point observation is given below:*

$$\mathbb{E}\left(N|\mathcal{Y}^{\mathrm{o}}\right) = \mathbb{E}(N)\ \eta((1-d))+\mathcal{Y}^{\mathrm{o}}\left(1-\beta_\gamma\right) \tag{12.34}$$

Notice that the first term in the r.h.s. of (12.34) represents the mean value of N times the nondetection probability. Roughly speaking, the second term represents the $\mathcal{Y}^{\mathrm{o}}$-probability that observations do not cause the clutter. In this connection, models with no clutter and fully detectable states are described below.

Corollary 12.3.13 *In the situation where* $d=1$ *and* $\nu=0$ *we have*

$$\mathcal{X}_c=\mathcal{X}\quad\text{and}\quad\mathcal{Y}^{\mathrm{o}}=\mathcal{Y}=\sum_{1\leq i\leq N}\delta_{Y^i}$$

In addition, for any function $F\in\mathcal{B}\left(\mathcal{M}(E_1)\right)$*, we have the following almost sure integral representation formula*

$$\mathbb{E}\left(F\left(\mathcal{X}\right)|\mathcal{Y}\right)=\int F\left(m_N(x)\right)\prod_{1\leq i\leq N}\Psi_{g(Y^i,\bullet)}(\eta)(dx^i)$$

For any function $f\in\mathcal{B}(E_1\cup\{c\})$ *we also have the almost sure integral representation formula*

$$\mathbb{E}\left(\mathcal{X}_c(f)\mid\mathcal{Y}\right)=\mathbb{E}\left(\mathcal{X}(f)\mid\mathcal{Y}\right)=\int\mathcal{Y}(dy)\ \Psi_{g(y,\bullet)}(\eta)(f)$$

12.4 Association tree based measures

This section is concerned with association tree based measures and their mean field approximations. The central idea behind these filtering association models is to solve the data association problem; that is, to find the right sequences of observations delivered by every target track.

In the first section, Section 12.4.1, we design a new class of evolution equations in the set of measures on finite association trees. This class of models has the same form as the ones discussed in Section 12.2.1.

Then, we examine two situations:

Firstly, for a given data association trajectory, we assume that the optimal single target tracking problem can be solved using Kalman filters, or by some auxiliary particle filter. In this context, the central problem is to find a judicious way of reducing the set of all possible associations to a reasonable finite number, with high likelihood value. In Section 12.4.2, we design a mean field solution to this problem.

In more general situations, even given the exact sequence of observations of a given target trajectory, the optimal filtering problem associated with this data cannot be computed explicitly. In Section 12.4.3, we couple the data association mean field model discussed in Section 12.4.2 with mean field type particle filters.

To the best of our knowledge, these mean field particle approximations of association tree based measures are one of the most performant algorithms, for solving multiple target tracking problems. We refer to the series of articles [145, 476, 477, 478] for numerical experiments and comparisons between these mean field models.

12.4.1 Nonlinear evolution equations

We let $(\mathcal{A}_n)_{n\geq 0}$ be a sequence of finite sets equipped with some finite positive measures $(\nu_n)_{n\geq 0}$. We let $\eta^{(a)}$, $Q^{(a)}$, and $f^{(a)}$ be some collection of measures, integral operators, and measurable functions, on some state spaces, indexed by the parameter a in some finite set $\mathcal{A}$.

To clarify the presentation, for any measure ν on $\mathcal{A}$, we set

$$\eta^{(\nu)} := \int \nu(da)\ \eta^{(a)} \tag{12.35}$$

$$Q^{(\nu)} := \int \nu(da)\ Q^{(a)} \quad \text{and} \quad f^{(\nu)} := \int \nu(da)\ f^{(a)} \tag{12.36}$$

We return to the general measure valued model (γ_n, η_n) defined in Section 12.2.1. We further assume that the initial distribution γ_0 and the integral operators Q_{n+1,γ_n} in (12.11) have the following form

$$\gamma_0 = \eta_0^{(\nu_0)} \quad \text{and} \quad Q_{n+1,\gamma_n} = Q_{n+1,\gamma_n}^{(\nu_{n+1})}$$

for some collection of probability measures $\eta_0^{(a)}$, and positive and bounded integral operators $Q_{n+1,\gamma_n}^{(a)}$, indexed by $a \in \mathcal{A}_{n+1}$. In this situation, we have

$$\gamma_0(1) = \nu_0(1) \quad \text{and} \quad \eta_0 = \eta_0^{(A_0)} \quad \text{with} \quad A_0(da) := \nu_0(da)/\nu_0(1)$$

We also assume that the following property is met

$$G_{n,\gamma}^{(a)} := Q_{n+1,\gamma}^{(a)}(1) \propto G_n^{(a)} \quad \text{and} \quad Q_{n+1,\gamma}^{(a)}(f)/Q_{n+1,\gamma}^{(a)}(1) := M_{n+1}^{(a)}(f) \tag{12.37}$$

for some function $G_n^{(a)}$ on E_n, and some Markov transitions $M_{n+1}^{(a)}$ from E_n into E_{n+1} whose values do not depend on the measures γ.

Example 12.4.1 *We illustrate these rather abstract conditions in the context of the multiple target tracking equation presented in (12.23).*

In this situation, it is convenient to add a pair of virtual observation states c, c' to E_n^Y. Using this notation, the above conditions are satisfied with the finite sets $\mathcal{A}_{n+1}$ and their counting measures ν_{n+1} defined below

$$\mathcal{A}_{n+1} = \left\{Y_n^i, 1 \leq i \leq N_n^Y\right\} \cup \{c, c'\}$$

and

$$\nu_{n+1} = \mathcal{Y}_n^{\circ} + \delta_c + \delta_{c'} \in \mathcal{M}(\mathcal{A}_{n+1})$$

Using (12.23) and (12.24), we check that (12.37) is met with a couple of potential functions and Markov transitions defined by

$$(G_n^{(a)}, M_{n+1}^{(a)}) = \begin{cases} (r_n d_n g_n(\boldsymbol{.}, a)\ ,\ M_{n+1}) & \text{for} \quad a \notin \{c, c'\} \\ (r_n(1-d_n)\ ,\ M_{n+1}) & \text{for} \quad a = c \\ (1\ ,\ \overline{\mu}_{n+1}) & \text{for} \quad a = c' \end{cases}$$

In this case, we observe that

$$Q_{n+1,\gamma_n}^{(a)}(x_n, \boldsymbol{.}) = G_{n,\gamma_n}^{(a)}(x_n)\ M_{n+1}^{(a)}(x_n, \boldsymbol{.})$$

with the potential function $G_{n,\gamma_n}^{(a)}$ defined below

$$G_{n,\gamma_n}^{(a)} / G_n^{(a)} = \begin{cases} [h_n(a) + \gamma_n(d_n g_n(\boldsymbol{.}, a))]^{-1} & \text{for} \quad a \notin \{c, c'\} \\ 1 & \text{for} \quad a = c \\ \mu_{n+1}(1)/\gamma_n(1) & \text{for} \quad a = c' \end{cases} \tag{12.38}$$

Definition 12.4.2 *We consider the collection of probability measures $\eta_n^{(\mathbf{a_n})} \in \mathcal{P}(E_n)$, indexed by sequences of parameters*

$$\mathbf{a_n} = (a_0, \ldots, a_n) \in \boldsymbol{\mathcal{A}_n} := (\mathcal{A}_0 \times \ldots \times \mathcal{A}_n)$$

and defined by the following equations

$$\eta_n^{(\mathbf{a_n})} \quad = \quad \left(\Phi_n^{(a_n)} \circ \ldots \circ \Phi_1^{(a_1)}\right)\left(\eta_0^{(a_0)}\right) \tag{12.39}$$

with the mappings $\Phi_n^{(a)} : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$, indexed by $a \in \mathcal{A}_n$, and defined by the updating-prediction Feynman-Kac transformation

$$\Phi_n^{(a)}(\eta) = \Psi_{G_{n-1}^{(a)}}(\eta)\, M_n^{(a)}$$

Given some $\mathbf{a} = (a_0, \ldots, a_n) \in \boldsymbol{\mathcal{A}_n}$, we set $|\mathbf{a}| = n$. In this situation, we simplify the notation, and write

$$\eta^{(\mathbf{a})} := \eta_{|\mathbf{a}|}^{(\mathbf{a}_{|\mathbf{a}|})} = \eta_n^{(\mathbf{a_n})} \tag{12.40}$$

the measure defined in (12.39).

Definition 12.4.3 *We let Ω_{n+1} be the mapping*

$$\Omega_{n+1} \ : \ (m, A) \in (]0, \infty[\times \mathcal{P}(\boldsymbol{\mathcal{A}_n})) \mapsto \Omega_{n+1}(m, A) \in \mathcal{P}(\boldsymbol{\mathcal{A}_{n+1}})$$

defined by the following formula

$$\Omega_{n+1}(m, A) = \Psi_{\mathcal{G}_{m,A}}(A \otimes \nu_{n+1}) \tag{12.41}$$

with the Boltzmann-Gibbs transformations $\Psi_{\mathcal{G}_{m,A}}$ from $\mathcal{P}(\boldsymbol{\mathcal{A}_n} \times \mathbb{R}_+)$, into itself defined in (7), associated with the potential function

$$\mathcal{G}_{m,A}(a, b) = \eta_n^{(a)}\left(G_{n,m\eta^{(A)}}^{(b)}\right) \qquad \text{with} \quad \eta^{(A)} \text{ given by (12.36) and (12.40).}$$

Definition 12.4.4 *We consider the collection of integral operators $\mathcal{Q}_{n+1,B}$ from $\boldsymbol{\mathcal{A}_n}$ into $\boldsymbol{\mathcal{A}_{n+1}}$, indexed by $B \in \mathcal{M}_+(\boldsymbol{\mathcal{A}_n})$, and defined, for any $a \in \boldsymbol{\mathcal{A}_n}$ by*

$$\mathcal{Q}_{n+1,B}(a, d(a',b)) := [\delta_a \otimes \nu_{n+1}]\,(d(a',b)) \quad \mathcal{G}_{B(1),\overline{B}}(a',b) \tag{12.42}$$

with the normalized distribution

$$\overline{B} = B/B(1) \in \mathcal{P}(\boldsymbol{\mathcal{A}_n})$$

In the above display formulae, $d(a',b) = da' \times db$ stands for an infinitesimal neighborhood of the point $(a',b) \in \boldsymbol{\mathcal{A}_{n+1}} := (\boldsymbol{\mathcal{A}_n} \times \mathcal{A}_{n+1})$.

Proposition 12.4.5 *The solution of (12.2.1) has the form*

$$\eta_n = \eta_n^{(A_n)} = \eta^{(A_n)}$$

with the sequence of association measures $A_n \in \mathcal{P}(\boldsymbol{\mathcal{A}_n})$ defined, for any $F \in \mathcal{B}_b(\boldsymbol{\mathcal{A}_{n+1}})$, by the evolution equations

$$A_{n+1}(F) = \Omega_{n+1}\,(\gamma_n(1), A_n)\,(F) := \frac{A_n \mathcal{Q}_{n+1,\gamma_n(1)A_n}(F)}{A_n \mathcal{Q}_{n+1,\gamma_n(1)A_n}(1)}$$

In addition, the flow of unnormalized measures $(B_n)_{n\geq 0} \in \mathcal{M}_+(\boldsymbol{\mathcal{A}_n})$ defined by $B_n := \gamma_n(1) \times A_n$ satisfies the same type of equation as in (12.11); that is, we have that

$$B_{n+1} = B_n \mathcal{Q}_{n+1,B_n} \tag{12.43}$$

Proof:
We check the first assertion using an inductive proof on the time parameter.

For $n = 0$, we have set $\eta_0 = \eta^{(A_0)}$, so that the assertion is met at the time $n = 0$. We further assume that $\eta_n = \eta_n^{(A_n)}$, for some $A_n \in \mathcal{P}(\boldsymbol{\mathcal{A}_n})$.

Using (12.11), we find that it is simply based on the fact that

$$\begin{aligned} \eta_{n+1} &\propto \eta^{(A_n)} Q^{(\nu_{n+1})}_{n+1,\gamma_n(1)\eta^{(A_n)}} \\ &= \int [A_n \otimes \nu_{n+1}](d(a,b))\ \eta_n^{(a)} Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}} \end{aligned}$$

Using (12.37), we have that

$$Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}}(1) = G^{(b)}_{n,\gamma_n(1)\eta^{(A_n)}} \propto G_n^{(b)}$$

and

$$Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}}(f)/Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}}(1) = M^{(b)}_{n+1}(f)$$

for any $f \in \mathcal{B}_b(E_{n+1})$. This implies that

$$\eta_{n+1}^{(a,b)}(f) = \Phi_{n+1}^{(b)}\left(\eta_n^{(a)}\right)(f) = \frac{\eta_n^{(a)}\left(G_n^{(b)} M_{n+1}^{(b)}(f)\right)}{\eta_n^{(a)}\left(G_n^{(b)}\right)} = \frac{\eta_n^{(a)} Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}}(f)}{\eta_n^{(a)} Q^{(b)}_{n+1,\gamma_n(1)\eta^{(A_n)}}(1)}$$

from which we conclude that

$$\eta_{n+1} \propto \int \underbrace{[A_n \otimes \nu_{n+1}](d(a,b))\ \eta_n^{(a)}\left(G^{(b)}_{n,\gamma_n(1)\eta^{(A_n)}}\right)}_{=\Omega_{n+1}(\gamma_n(1),A_n)(d(a,b))}\ \eta_{n+1}^{(a,b)}$$

In terms of the second coordinate mapping Λ^2_{n+1} defined in (12.14), we have proved that

$$\eta_{n+1} = \Lambda^2_{n+1}\left(\gamma_n(1), \eta_n^{(A_n)})\right) = \eta_{n+1}^{(\Omega_{n+1}(\gamma_n(1),A_n))} = \eta_{n+1}^{(A_{n+1})}$$

This ends the proof of the first assertion.
On the other hand, we have that

$$\begin{aligned} A_n \mathcal{Q}_{n+1,\gamma_n(1)A_n}(F)(a) &= \int A_n(da)\ \nu_{n+1}(db)\ \mathcal{G}_{\gamma_n(1),A_n}(a,b)\ F(a,b) \\ &\propto \int A_{n+1}(d(a,b))\ F(a,b) \end{aligned}$$

This ends the proof of the proposition. ∎

Using Proposition 12.4.5, we readily prove the following theorem.

Theorem 12.4.6 *The solution of (12.2.1) has the form*

$$\gamma_n = \gamma_n(1) \times \eta^{(A_n)}$$

with the process $(\gamma_n(1), A_n) \in (\mathbb{R}_+ \times \mathcal{P}(\boldsymbol{\mathcal{A}}_n))$ *defined by*

$$\left\{ \begin{array}{rcl} A_{n+1} &=& \Psi_{\mathcal{G}_{\gamma_n(1),A_n}}(A_n \otimes \nu_{n+1}) \\ \gamma_{n+1}(1) &=& \gamma_n(1) \times [A_n \otimes \nu_{n+1}](\mathcal{G}_{\gamma_n(1),A_n}) \end{array} \right. \tag{12.44}$$

In the above display, $\mathcal{G}_{\gamma_n(1),A_n}$ *stands for the potential function presented in Definition 12.4.3.*

We end this section with some comments on the evolution Equation (12.44). The first equation is only defined in terms of the Boltzmann-Gibbs transformations $\Psi_{\mathcal{G}_{\gamma_n(1),A_n}}$. As a result, we cannot expect the measure valued equation to have some nice stability properties.

Loosely speaking, we can stabilize these equations adding some MCMC steps between the updating Boltzmann-Gibbs transformations. More formally, we can add, at every time step, an MCMC transition $\boldsymbol{M_{n+1}}$ on the set $\boldsymbol{\mathcal{A}}_{n+1}$ with invariant measure $\Psi_{\mathcal{G}_{\gamma_n(1),A_n}}$. The resulting equation, is now given by

$$A_{n+1} = \Psi_{\mathcal{G}_{\gamma_n(1),A_n}}(A_n \otimes \nu_{n+1})\, \boldsymbol{M_{n+1}}$$

12.4.2 Mean field particle model

We further assume that $\eta_n^{(a)}\left(G^{(b)}_{n,\gamma_n(1)\eta^{(A_n)}}\right)$ are explicitly known for any sequence of parameters $((a,A_n),b) \in \left(\boldsymbol{\mathcal{A}}^{\mathbf{2}}_{\boldsymbol{n}} \times \mathcal{A}_{n+1}\right)$.

This rather strong condition is satisfied for the multiple target tracking model discussed above as long as the quantities

$$\eta_n^{(a_0,y_0,\ldots,y_{n-1})}(r_n d_n g_n(.,y_n)) \qquad \eta_n^{(a_0,y_0,\ldots,y_{n-1})}(r_n(1-d_n))$$

and

$$\eta_n^{(a_0,y_0,\ldots,y_{n-1})}(d_n g_n(.,y_n))$$

are explicitly known. This condition is clearly met for linear Gaussian target evolution and observation sensors, as long as the survival and detection probabilities, s_n and d_n, are state independent, and the spontaneous birth $\overline{\mu}_n$ as well as the spawned targets branching rates b_n are Gaussian mixtures.

In this situation, the collections of measures $\eta_n^{(a_0,y_0,\ldots,y_{n-1})}$ are Gaussian distributions and the Equation (12.39) coincides with the traditional updating-prediction transitions of the discrete generation Kalman-Bucy filter.

We let $A_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_0^i}$ be the empirical measure associated with N independent and identically distributed random variables $(\mathrm{a}_0^i)_{1\leq i\leq N}$ with common distribution A_0. By construction, we have

$$\eta_0^N := \int A_0^N(da)\ \eta_0^{(a)} \ \simeq_{N\uparrow\infty}\ \eta_0$$

We further assume that $\gamma_0(1)$ is known and we set $\gamma_0^N = \gamma_0(1)\ \eta_0^N$,

$$\gamma_1^N(1) = \gamma_0^N(1)\ \eta_0^N(G_{0,\gamma_0^N})\ , \quad \text{and} \quad \eta_1^N := \int A_1^N(da)\ \eta_1^{(a)}$$

with the occupation measure $A_1^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_1^i}$ associated with N conditionally independent and identically distributed random variables $\mathrm{a}_1^i := (a_{0,1}^i, a_{1,1}^i)$ with common law $\Omega_1\left(\gamma_0^N(1), A_0^N\right)$. By construction, we also have

$$\eta_1^N \ \simeq_{N\uparrow\infty}\ \int \Omega_1\left(\gamma_0^N(1), A_0^N\right)(da)\ \eta_1^{(a)} = \Lambda_1^2\left(\gamma_0^N(1), \eta_0^N\right)$$

Iterating this procedure, we define by induction a sequence of N-particle approximation measures

$$\gamma_n^N(1) = \gamma_{n-1}^N(1)\ \eta_{n-1}^N(G_{n-1,\gamma_{n-1}^N}) \quad \text{and} \quad \eta_n^N := \int A_n^N(da)\ \eta_n^{(a)}$$

with the occupation measure $A_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_n^i}$ associated with N conditionally independent and identically distributed random variables

$$\mathrm{a}_n^i := (a_{0,n}^i, a_{1,n}^i, \ldots, a_{n,n}^i)$$

with common law $\Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N\right)$. Arguing as above, we find that

$$\eta_n^N \ \simeq_{N\uparrow\infty}\ \int \Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N\right)(da)\ \eta_n^{(a)} = \Lambda_n^2\left(\gamma_{n-1}^N(1), \eta_{n-1}^N\right)$$

12.4.3 Mixed particle association models

We consider the association mapping

$$\boldsymbol{\Omega_{n+1}}\ :\ (m,A,\eta) \in \left(]0,\infty[\times\boldsymbol{\mathcal{A}_n}\times\mathcal{P}(E_n)^{\boldsymbol{\mathcal{A}_n}}\right) \mapsto \Omega_{n+1}(m,A,\eta) \in \mathcal{P}(\boldsymbol{\mathcal{A}_{n+1}})$$

defined for any $(m,A) \in (]0,\infty[\times\boldsymbol{\mathcal{A}_n})$, and any mapping

$$\eta : a \in \mathrm{Supp}(A) \mapsto \eta^{(a)} \in Pa(E_n)$$

by the following equation

$$\boldsymbol{\Omega_{n+1}}\,(m,A,\eta)\,(d(a,b)) \propto A(da)\ \nu_{n+1}(db)\ \eta^{(a)}\left(G_{n,m\int A(da)\ \eta^{(a)}}^{(b)}\right)$$

By construction, for any discrete measure $A \in \mathcal{P}(\boldsymbol{\mathcal{A}_{n-1}})$, and any mapping $a \in \mathrm{Supp}(A) \mapsto \eta^{(a)} \in \mathcal{P}(E_{n-1})$, we have the formula

$$\Lambda_n^2\left(m, \int A(da)\ \eta^{(a)})\right) = \int \boldsymbol{\Omega_n}\left(m, A, \eta^{(\cdot)}\right)(d(a,b))\ \ \Phi_n^{(b)}\left(\eta^{(a)}\right)$$

We also mention that the updating-prediction transformation defined in (12.39) can be rewritten in terms of nonlinear transport equations

$$\Phi_n^{(a)}(\eta) = \Psi_{G_{n-1}}^{(a)}(\eta)\, M_n^{(a)} = \eta \mathcal{K}_{n,\eta}^{(a)} \quad \text{with} \quad \mathcal{K}_{n,\eta}^{(a)} = \mathcal{S}_{n-1,\eta}^{(a)} M_n^{(a)} \tag{12.45}$$

In the above displayed formula, $\mathcal{S}_{n-1,\eta}^{(a)}$ stands for some updating Markov transition, from E_{n-1} into itself, satisfying the compatibility condition

$$\eta \mathcal{S}_{n-1,\eta}^{(a)} = \Psi_{G_{n-1}}^{(a)}(\eta)$$

We let $A_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_0^i}$ be the empirical measure associated with N independent and identically distributed random variables $(\mathrm{a}_0^i)_{1\leq i\leq N}$ with common distribution A_0. We set

$$\eta_0^N := \int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a},N')}$$

with the empirical measure $\eta_0^{(\mathrm{a},N')} = \frac{1}{N'}\sum_{i=1}^{N'} \delta_{\xi_0^{[\mathrm{a},j]}}$ associated with N' random variables $\xi_0^{[\mathrm{a}]} = \left(\xi_0^{[\mathrm{a},j]}\right)_{1\leq j\leq N'}$, with common law $\eta_0^{(\mathrm{a})}$. We further assume that $\gamma_0(1)$ is known, and we set

$$\gamma_0^N := \gamma_0(1)\ \eta_0^N \quad \text{and} \quad \gamma_1^N(1) := \gamma_0^N(1)\ \eta_0^N(G_{0,\gamma_0^N})$$

By construction, we have

$$\int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a},N')} \simeq_{N'\uparrow\infty} \int A_0^N(d\mathrm{a})\ \eta_0^{(\mathrm{a})} \Rightarrow \eta_0^N \simeq_{N,N'\uparrow\infty} \eta_0$$

Using (12.45), for any $\mathrm{a}_1 = (a_0, a_1) \in \boldsymbol{\mathcal{A}_1}$ we find that

$$\Phi_1^{(a_1)}\left(\eta_0^{(a_0,N')}\right) = \eta_0^{(a_0,N')} \mathcal{K}_{n,\eta_0^{(a_0,N')}}^{(a_1)}$$

We let $A_1^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_1^i}$ be the occupation measure associated with N conditionally i.i.d. r.v. $\mathrm{a}_1^i := (a_{0,1}^i, a_{1,1}^i)$ with common law

$$\boldsymbol{\Omega_1}\left(\gamma_0^N(1), A_0^N, \eta_0^{(\cdot,N')}\right)$$

In the above displayed formula $\eta_0^{(\cdot,N')}$ stands for the mapping

$$a_0 \in \mathcal{A}_0 \mapsto \eta_0^{(a_0,N')} \in \mathcal{P}(E_0)$$

We consider a sequence of conditionally independent random variables $\xi_1^{[a_0,a_1,j]}$ with distribution $\mathcal{K}_{n,\eta_0^{(a_0,N')}}^{(a_1)}\left(\xi_0^{[a_0,j]}, .\right)$, with $1 \leq j \leq N'$, and we set

$$\eta_1^{((a_0,a_1),N')} = \frac{1}{N'}\sum_{i=1}^{N'} \delta_{\xi_1^{[(a_0,a_1),j]}} \quad \text{and} \quad \eta_1^N := \int A_1^N(da)\ \eta_1^{(\mathrm{a},N')}$$

Arguing as before, we find that

$$\begin{aligned} \eta_1^N &\simeq_{N\uparrow\infty} \int \boldsymbol{\Omega_1}\left(\gamma_0^N(1), A_0^N, \eta_0^{(\cdot,N')}\right)(d(a_0,a_1))\ \Phi_1^{(a_1)}\left(\eta_0^{(a_0,N')}\right) \\ &= \Lambda_1^2\left(\gamma_0^N(1), \eta_0^N\right) \end{aligned}$$

Iterating this procedure, we define by induction a sequence of N-particle approximation measures

$$\gamma_n^N(1) = \gamma_{n-1}^N(1)\ \eta_{n-1}^N(G_{n-1,\gamma_{n-1}^N}) \quad \text{and} \quad \eta_n^N := \int A_n^N(da)\ \eta_n^{(a,N')}$$

with the occupation measure $A_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathrm{a}_n^i}$ associated with N conditionally independent and identically distributed random variables

$$\mathrm{a}_n^i := (a_{0,n}^i, a_{1,n}^i, \ldots, a_{n,n}^i)$$

with common law

$$\Omega_n\left(\gamma_{n-1}^N(1), A_{n-1}^N, \eta_{n-1}^{(\cdot,N')}\right)$$

Arguing as above, we find that

$$\begin{aligned}\eta_n^N &\simeq_{N\uparrow\infty} \int \mathbf{\Omega_n}\left(\gamma_{n-1}^N(1), A_{n-1}^N, \eta_{n-1}^{(\cdot,N')}\right)(d(a,b))\ \Phi_n^{(b)}\left(\eta_{n-1}^{(a,N')}\right)\\ &= \Lambda_n^2\left(\gamma_{n-1}^N(1), \eta_{n-1}^N\right)\end{aligned}$$

As before, the N-particle occupation measures A_n^N converge as N tends to ∞ to the association probability measures A_n.

Part IV

Solution of the exercises

Solution to exercise 4.7.1:
Choosing $R = R_\epsilon := 2/(1-\epsilon)$, we find that

$$\left(1 - \left(\epsilon + \frac{1}{R}\right)\right) = \left(1 - \left(\epsilon + \frac{1-\epsilon}{2}\right)\right) = 1 - \frac{1+\epsilon}{2} = \frac{1-\epsilon}{2}$$

and

$$1 - \frac{1}{1+2\rho R} = 1 - \frac{(1-\epsilon)}{(1-\epsilon)+4\rho} = \frac{4\rho}{(1-\epsilon)+4\rho}$$

This implies that

$$W(x) \wedge W(y) \geq R_\epsilon$$
$$\Downarrow$$
$$\forall \rho \in]0,1] \qquad \frac{\|M(x,.) - M(y,.)\|_{V_\rho}}{1 + V_\rho(x) + V_\rho(y)} \leq 1 - \frac{1}{2}\,\frac{4\rho(1-\epsilon)}{(1-\epsilon)+4\rho} < 1$$

In much the same way, when $W(x) \vee W(y) \leq 2/(1-\epsilon)$ we have

$$\frac{\|M(x,.) - M(y,.)\|_{V_\rho}}{1 + V_\rho(x) + V_\rho(y)} \leq 1 - \left(\alpha_\epsilon - \frac{8\rho}{1-\epsilon}\right) < 1$$

with $\alpha_\epsilon := 1 - \beta^{(R_\epsilon)}(M)$, as soon as

$$\rho < \alpha_\epsilon \delta/8 \quad \text{with} \quad \delta := (1-\epsilon)$$

If we set $u := 4\rho/\delta$, then we have

$$\frac{1}{2}\,\frac{4\rho(1-\epsilon)}{(1-\epsilon)+4\rho} = \frac{\delta}{2}\,\left(1 - \frac{1}{1+u}\right) := g(u)$$

and

$$\left(\alpha_\epsilon - \frac{8\rho}{1-\epsilon}\right) = (\alpha_\epsilon - 2u) := h(u)$$

On the interval $u \in [0, \alpha_\epsilon/2]$ (so that $\rho < \alpha_\epsilon\delta/8$) the function g is increasing from $g(0) = 0$ to $g(\alpha_\epsilon/2) = \frac{\delta\alpha_\epsilon}{2+\alpha_\epsilon} < 1$, while the function h is decreasing from $h(0) = \alpha_\epsilon$ to $h(\alpha_\epsilon/2) = 0$. These two functions intersects at some point u such that

$$(1+u)\ (2u - \alpha_\epsilon) + u\delta/2 = 0$$

In other words, if we set

$$a := \frac{1}{2}\left(1 - b + \frac{\delta}{4}\right) \leq \frac{1}{2} \quad \text{with} \quad b := \frac{\alpha_\epsilon}{2}$$

we need to solve the equation

$$u^2 + 2ua - b = (u-a)^2 - \left[a^2 + b\right] = 0$$

with $u \in [0, b]$. This implies that

$$0 < u = \sqrt{a^2 + b} - a \leq b$$

The r.h.s. inequality is checked using the fact that

$$\begin{aligned} \sqrt{a^2+b} - a \leq b \quad &\Leftrightarrow \quad a^2 + b \leq a^2 + b^2 + 2ab \\ &\Leftrightarrow \quad b \leq b\ (b + 2a) = b\ \left(1 + \frac{\delta}{4}\right) \end{aligned}$$

Using a Taylor expansion, for any $v \geq 0$ we have

$$\sqrt{1+v} = 1 + \frac{v}{2\sqrt{1+v\tau_v}} \geq 1 + \frac{v}{2\sqrt{1+v}}$$

for some $\tau_v \in [0,1]$. If we set $v = b/a^2$ we find that

$$\begin{aligned} u = \sqrt{a^2+b} - a \geq \tfrac{b}{2\sqrt{a^2+b}} \\ \Rightarrow \quad g(u) = h(u) \quad &\geq \quad g\left(\tfrac{b}{2\sqrt{a^2+b}}\right) \\ &= \quad \tfrac{\delta}{2} \, \tfrac{b}{2\sqrt{a^2+b}} \, \tfrac{1}{1+\frac{b}{2\sqrt{a^2+b}}} \\ &= \quad \tfrac{\delta b}{2} \, \tfrac{1}{b+2\sqrt{a^2+b}} \geq \tfrac{\delta b}{1+2\sqrt{3}} \end{aligned}$$

The r.h.s. estimate comes from the fact that

$$a \vee b \leq 1/2 \;\Rightarrow\; b + \sqrt{4a^2+4b} \leq 1/2 + \sqrt{1+2} = (1+2\sqrt{3})/2$$

Choosing $\rho = u\delta/4$, we conclude that

$$\beta_{V_\rho}(M) = \sup_{x,y} \frac{\|M(x,.) - M(y,.)\|_{V_\rho}}{1 + V_\rho(x) + V_\rho(y)} \leq 1 - g(u) \leq 1 - \frac{(1-\epsilon)(1-\beta^{(R_\epsilon)}(M))}{2(1+2\sqrt{3})}$$

This ends the proof of the exercise.

■

Solution to exercise 4.7.2:

Since M has positive entries, by theorem **??** all the entries of γ are positive. We let $\gamma = [\gamma(1), \gamma(2), \gamma(3)]$. We want to solve the equation

$$[\gamma(1), \gamma(2), \gamma(3)] \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = [\gamma(1), \gamma(2), \gamma(3)]$$

In other words, we have

$$\left\{ \begin{array}{rcl} \gamma(1)\ [1-(p_{12}+p_{13})] + \gamma(2)\ p_{21} + \gamma(3) p_{31} &=& \gamma(1) \\ \gamma(1)\ p_{12} + \gamma(2)\ [1-(p_{21}+p_{23})] + \gamma(3) p_{32} &=& \gamma(2) \\ \gamma(1)\ p_{13} + \gamma(2)\ p_{23} + \gamma(3)\ [1-(p_{31}+p_{32})] &=& \gamma(3) \end{array} \right.$$

which is equivalent to

$$\left\{ \begin{array}{rcl} \gamma(2)\ p_{21} + \gamma(3)\ p_{31} &=& \gamma(1)\ [p_{12}+p_{13}] \\ \gamma(1)\ p_{12} + \gamma(3)\ p_{32} &=& \gamma(2)\ [p_{21}+p_{23}] \\ \gamma(1)\ p_{13} + \gamma(2)\ p_{23} &=& \gamma(3)\ [p_{31}+p_{32}] \end{array} \right.$$

This yields the system

$$\left\{ \begin{array}{rcl} \frac{\gamma(2)}{\gamma(1)}\ p_{21} + \frac{\gamma(3)}{\gamma(1)}\ p_{31} &=& [p_{12}+p_{13}] \\ p_{12} + \frac{\gamma(3)}{\gamma(1)}\ p_{32} &=& \frac{\gamma(2)}{\gamma(1)}\ [p_{21}+p_{23}] \end{array} \right.$$

This shows that

$$\left\{ \begin{array}{rcl} \frac{\gamma(2)}{\gamma(1)}\ p_{21} + \frac{\gamma(3)}{\gamma(1)}\ p_{31} &=& [p_{12}+p_{13}] \\ \frac{\gamma(2)}{\gamma(1)}\ [p_{21}+p_{23}] - \frac{\gamma(3)}{\gamma(1)}\ p_{32} &=& p_{12} \end{array} \right.$$

Multiplying the first line by $[p_{21}+p_{23}]$ and the second one by p_{21}, we find that

$$\left\{\begin{array}{rcl} \frac{\gamma(2)}{\gamma(1)}\ p_{21}\left[p_{21}+p_{23}\right] & +\frac{\gamma(3)}{\gamma(1)}\ p_{31}\left[p_{21}+p_{23}\right] & = \left[p_{21}+p_{23}\right]\ \left[p_{12}+p_{13}\right] \\ \frac{\gamma(2)}{\gamma(1)}\ p_{21}\left[p_{21}+p_{23}\right] & -\frac{\gamma(3)}{\gamma(1)}\ p_{21}p_{32} & = p_{12}p_{21} \end{array}\right.$$

Then we subtract the two lines to check that

$$\begin{aligned} \frac{\gamma(3)}{\gamma(1)}\ \left(p_{31}\left[p_{21}+p_{23}\right]+p_{21}p_{32}\right) &= \left[p_{21}+p_{23}\right]\ \left[p_{12}+p_{13}\right]-p_{12}p_{21} \\ &= p_{21}p_{13}+p_{23}\left[p_{12}+p_{13}\right] \end{aligned}$$

This implies that

$$\frac{\gamma(3)}{\gamma(1)}=\frac{p_{21}p_{13}+p_{23}\left[p_{12}+p_{13}\right]}{p_{31}\left[p_{21}+p_{23}\right]+p_{21}p_{32}}$$

In much the same way, multiplying the first line by p_{32} and the second one by p_{31}, we find that

$$\left\{\begin{array}{rcl} \frac{\gamma(2)}{\gamma(1)}\ p_{21}p_{32} & +\frac{\gamma(3)}{\gamma(1)}\ p_{31}p_{32} & = p_{32}\left[p_{12}+p_{13}\right] \\ \frac{\gamma(2)}{\gamma(1)}\ p_{31}\left[p_{21}+p_{23}\right] & -\frac{\gamma(3)}{\gamma(1)}\ p_{31}p_{32} & = p_{12}p_{31} \end{array}\right.$$

Adding the two lines we find that

$$\frac{\gamma(2)}{\gamma(1)}\left(p_{31}\left[p_{21}+p_{23}\right]+p_{21}p_{32}\right)=p_{32}\left[p_{12}+p_{13}\right]+p_{12}p_{31}$$

from which we conclude that

$$\frac{\gamma(2)}{\gamma(1)}=\frac{p_{32}\left[p_{12}+p_{13}\right]+p_{12}p_{31}}{p_{31}\left[p_{21}+p_{23}\right]+p_{21}p_{32}}$$

and

$$\frac{\gamma(3)}{\gamma(2)}=\frac{\gamma(3)}{\gamma(1)}\times\frac{\gamma(1)}{\gamma(2)}=\frac{p_{21}p_{13}+p_{23}\left[p_{12}+p_{13}\right]}{p_{32}\left[p_{12}+p_{13}\right]+p_{12}p_{31}}$$

We conclude that

$$\begin{aligned} \gamma(1) &\propto p_{31}\left[p_{21}+p_{23}\right]+p_{21}p_{32} \\ &= p_{31}p_{21}+p_{23}p_{31}+p_{32}p_{21}=\prod_{(i,j)\in g_1}p_{i,j}+\prod_{(i,j)\in g_3}p_{i,j}+\prod_{(i,j)\in g_3}p_{i,j} \end{aligned}$$

with the 1-graphs $\{g_1,g_2,g_3\}$ defined on page 85. and

$$\begin{aligned} \gamma(2) &\propto p_{32}\left[p_{12}+p_{13}\right]+p_{12}p_{31} \\ \gamma(3) &\propto p_{21}p_{13}+p_{23}\left[p_{12}+p_{13}\right] \end{aligned}$$

This ends the proof of the exercise.

■

Solution to exercise 4.7.3:

For any $g\in\mathcal{G}(x)$ and $x'\not\in x$, the set $h=g\cup\{(x,x')\}$ is a directed graph on S with a single loop at the state x'. We let $\mathcal{L}(x')$ the set of these graphs. We clearly have that

$$\mathcal{L}(x')=\cup_{x\neq x'}\ \left(\mathcal{G}(x)\cup\{(x,x')\}\right)=\cup_{x\neq x'}\ \left(\mathcal{G}(x')\cup\{(x',x')\}\right)$$

We set

$$M(g):=\prod_{(u,v)\in g}M(u,v)$$

In this notation, we have that

$$\begin{aligned}
\sum_{x\ :\ x\neq x'} \gamma(x)M(x,x') &= \sum_{x\ :\ x\neq x'} \left[\sum_{g\in\mathcal{G}(x)} M(g)\right] M(x,x') \\
&= \sum_{x\ :\ x\neq x'} \sum_{g\in\mathcal{G}(x)} M(g\cup\{(x,x')\}) \\
&= \sum_{h\in\mathcal{L}(x')} M(h) \\
&= \sum_{x\ :\ x\neq x'} \sum_{g\in\mathcal{G}(x')} M(g\cup\{(x',x)\}) \\
&= \sum_{x\ :\ x\neq x'} \gamma(x')M(x',x) = \gamma(x')\ \left(1-M(x',x')\right)
\end{aligned}$$

The end of the proof of the exercise is now clear. ∎

Solution to exercise 4.7.4:

$$P(\lambda) = \mathrm{Det}\begin{pmatrix} p_{11}-\lambda & p_{12} & p_{13} \\ p_{21} & p_{22}-\lambda & p_{23} \\ p_{31} & p_{32} & p_{33}-\lambda \end{pmatrix}$$

$$\begin{aligned}
&= (p_{11}-\lambda)\,\mathrm{Det}\begin{pmatrix} p_{22}-\lambda & p_{23} \\ p_{32} & p_{33}-\lambda \end{pmatrix} \\
&\qquad -p_{12}\,\mathrm{Det}\begin{pmatrix} p_{21} & p_{23} \\ p_{31} & p_{33}-\lambda \end{pmatrix} + p_{1,3}\,\mathrm{Det}\begin{pmatrix} p_{21} & p_{22}-\lambda \\ p_{31} & p_{32} \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
P(\lambda) &= (p_{11}-\lambda)\,[(p_{22}-\lambda)(p_{33}-\lambda)-p_{23}p_{32}] \\
&\qquad -p_{12}\,[p_{21}(p_{33}-\lambda)-p_{23}p_{31}] \\
&\qquad +p_{1,3}\,[p_{21}p_{32}-p_{31}(p_{22}-\lambda)] \\
&= -\lambda^3+\lambda^2 A+\lambda B+C
\end{aligned}$$

with

$$\begin{aligned}
A &= p_{11}+p_{22}+p_{33} \\
B &= p_{23}p_{32}+p_{12}p_{21}+p_{13}p_{31}-(p_{11}p_{22}+p_{11}p_{33}+p_{22}p_{33}) \\
C &= 1-(A+B)
\end{aligned}$$

The last assertion comes from the fact that $P(1)=0$ so that $A+B+C=1$. We also have that

$$P(\lambda) = (1-\lambda)\ \left(\lambda^2+(1-A)\lambda+C\right)$$

and

$$\lambda^2+(1-A)\lambda+C = \left(\lambda+\left(\frac{1-A}{2}\right)\right)^2-\left(\left(\frac{1-A}{2}\right)^2-C\right)$$

We also notice that

$$\begin{aligned}
1-A &= 1-((1-p_{12}-p_{13})+(1-p_{21}-p_{23})+(1-p_{31}-p_{32})) \\
&= p_{12}+p_{21}+p_{13}+p_{31}+p_{32}+p_{23}-2
\end{aligned}$$

This yields that

$$-\frac{1-A}{2} = 1 - (q_{12} + q_{13} + q_{23})$$

with

$$q_{i,j} = (p_{ij} + p_{ji})/2$$

from which we prove that

$$\begin{aligned}\left(\frac{1-A}{2}\right)^2 &= 1 + (q_{12} + q_{13} + q_{23})^2 - 2(q_{12} + q_{13} + q_{23}) \\ &= 1 + (q_{12}^2 + q_{13}^2 + q_{23}^2) + 2\,(q_{12}q_{13} + q_{12}q_{23} + q_{13}q_{23}) \\ &\qquad -2(q_{12} + q_{13} + q_{23})\end{aligned}$$

On the other hand, we have

$$\begin{aligned}p_{11}p_{22} &= (1 - p_{12} - p_{13})(1 - p_{21} - p_{23}) \\ &= 1 - (p_{12} + p_{13} + p_{21} + p_{23}) + (p_{12}p_{21} + p_{12}p_{23} + p_{13}p_{21} + p_{13}p_{23}) \\ p_{11}p_{33} &= (1 - p_{12} - p_{13})(1 - p_{31} - p_{32}) \\ &= 1 - (p_{12} + p_{13} + p_{31} + p_{32}) + (p_{12}p_{31} + p_{12}p_{32} + p_{13}p_{31} + p_{13}p_{32}) \\ p_{22}p_{33} &= (1 - p_{21} - p_{23})(1 - p_{31} - p_{32}) \\ &= 1 - (p_{21} + p_{23} + p_{31} + p_{32}) + (p_{21}p_{31} + p_{21}p_{32} + p_{23}p_{31} + p_{23}p_{32})\end{aligned}$$

from which we conclude that

$$\begin{aligned}B &= p_{23}p_{32} + p_{12}p_{21} + p_{13}p_{31} - [p_{11}p_{22} + p_{11}p_{33} + p_{22}p_{33}] \\ &= -3 + 4\ (q_{12} + q_{13} + q_{23}) - D\end{aligned}$$

with

$$\begin{aligned}D = (p_{12}p_{23} + p_{21}p_{32} + p_{12}p_{32}) + (p_{13}p_{21} + p_{21}p_{31} + p_{12}p_{31}) \\ + (p_{13}p_{23} + p_{23}p_{31} + p_{13}p_{32})\end{aligned}$$

We also have that

$$\begin{aligned}4q_{12}q_{23} &= (p_{12}p_{23} + p_{21}p_{32} + p_{12}p_{32}) + p_{21}p_{23} \\ 4q_{12}q_{13} &= (p_{21}p_{13} + p_{21}p_{31} + p_{12}p_{31}) + p_{12}p_{13} \\ 4q_{13}q_{23} &= (p_{13}p_{23} + p_{23}p_{31} + p_{13}p_{32}) + p_{31}p_{32}\end{aligned}$$

from which we prove that

$$D = 4\,(q_{12}q_{13} + q_{12}q_{23} + q_{13}q_{23}) - (p_{21}p_{23} + p_{12}p_{13} + p_{31}p_{32})$$

$$\left(\tfrac{1-A}{2}\right)^2 - C$$

$$= \left(\tfrac{1-A}{2}\right)^2 - 1 + (A + B)$$

$$= (q_{12}^2 + q_{13}^2 + q_{23}^2) + 2\,(q_{12}q_{13} + q_{12}q_{23} + q_{13}q_{23}) - 2(q_{12} + q_{13} + q_{23})$$

$$+ (3 - (p_{12} + p_{21} + p_{13} + p_{31} + p_{32} + p_{23}))$$

$$-3 + 4\ (q_{12} + q_{13} + q_{23}) - D$$

$$= (q_{12}^2 + q_{13}^2 + q_{23}^2) - 2\,(q_{12}q_{13} + q_{12}q_{23} + q_{13}q_{23})$$

$$+ [p_{21}p_{23} + p_{12}p_{13} + p_{31}p_{32}]$$

This implies that

$$\left(\frac{1-A}{2}\right)^2 - C = \Delta(q) + \delta(p)$$

with the parameters

$$\begin{aligned}\Delta(q) &= \frac{1}{2}\left[(q_{12}-q_{13})^2 + (q_{12}-q_{23})^2 + (q_{13}-q_{23})^2\right] \\ \delta(p) &= [p_{12}p_{13} - q_{12}q_{13}] + [p_{21}p_{23} - q_{21}q_{23}] + [p_{31}p_{32} - q_{31}q_{32}]\end{aligned}$$

This implies that

$$\begin{aligned}\lambda_2 &= (1-(q_{12}+q_{13}+q_{23})) + \sqrt{\Delta(q)+\delta(p)} \\ \lambda_3 &= (1-(q_{12}+q_{13}+q_{23})) - \sqrt{\Delta(q)+\delta(p)}\end{aligned}$$

with the convention $\sqrt{-a} = i\sqrt{a}$, for any $a \geq 0$. In the reversible case, we have $\delta(p) = 0$ and

$$\begin{aligned}\lambda_2 &= (1-(p_{12}+p_{13}+p_{23})) + \sqrt{\Delta(p)} \\ \lambda_3 &= (1-(p_{12}+p_{13}+_{23})) - \sqrt{\Delta(p)}\end{aligned}$$

We also check that

$$\lambda_2 \leq 1 \Leftrightarrow \frac{1}{2}\left[(p_{12}-p_{13})^2 + (p_{12}-p_{23})^2 + (p_{13}-p_{23})^2\right] \leq (p_{12}+p_{13}+p_{23})$$

Since

$$\begin{aligned}&\tfrac{1}{2}\left[(p_{12}-p_{13})^2 + (p_{12}-p_{23})^2 + (p_{13}-p_{23})^2\right] - (p_{12}+p_{13}+p_{23}) \\ &= -(p_{12}p_{13} + p_{12}p_{23} + p_{13}p_{23})\end{aligned}$$

we conclude that $\lambda_3 \leq \lambda_2 \leq \lambda_1$.

This ends the proof of the exercise.

■

Bibliography

[1] E. H. L. Aarts and J. H. M. Korst. *Simulated Annealing and Boltzmann Machines.* Wiley (1989).

[2] E. H. L. Aarts and J. K. Lenstra. *Local Search in Combinatorial Optimization.* Wiley (1989).

[3] Y. Achdou and O. Pironneau. Computational methods for option pricing. *SIAM, Frontiers in Applied Mathematics series* (2005).

[4] Y. Achdou, F. Camilli, and I. Capuzzo Dolcetta. Mean field games: convergence of a finite difference method. ArXiv:1207.2982v1 (2012).

[5] E. Aïdékon, B. Jaffuel. Survival of branching random walks with absorption. *Stochastic Processes and Their Applications*, vol. 121, pp. 1901–1937 (2011).

[6] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *EJP*, vol. 13, no. 34, pp. 1000–1034 (2008).

[7] D. Aldous and U. Vazirani. Go with the Winners Algorithms. In Proc. 35th Symp. Foundations of Computer Sci., pp. 492–501 (1994).

[8] K. Allen, C. Yau, and J. A. Noble. A recursive, stochastic vessel segmentation framework that robustly handles bifurcations. In Proceedings, Medical Image Understanding and Analysis, pp. 19–23 (2008).

[9] T.W. Anderson. An introduction to multivariate statistical analysis, 3rd edn. John Wiley and Sons, New York (2003).

[10] D. Andersson, B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics and Optimization*, vol. 63, no. 3, pp. 341–356 (2010).

[11] J. L. Anderson and S. L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, vol. 127, no. 12, pp. 2741–2758 (1999).

[12] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal Royal Statistical Society B.*, vol. 72, no. 3, pp. 269–342 (2010).

[13] C. Andrieu and A. Doucet. Particle filtering for partially observed Gaussian state space Models. *J. Royal Statistical Society* B, vol. 64, no. 4, pp. 827–836 (2002).

[14] F. Antonelli and A. Kohatsu-Higa. Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability*, vol. 12, no. 2, pp. 423–476 (2002).

[15] L. Arnold, L. Demetrius, and V.M. Gundlach. Evolutionary formalism for products of positive random matrices. *Ann. Appl. Probab.*, vol. 4, 859–901, (1994).

[16] J. Ashkin, E. Teller. Statistics of Two-Dimensional Lattices With Four Components. Phys. Rev., vol. 64, no. 5-6, pp. 178–184 (1943).

[17] D. Assaf, L. Goldstein, and E. Samuel-Cahn. An unexpected connection between branching processes and optimal stopping. *J. Appl. Prob.*, vol. 37, pp. 613–626 (2000).

[18] R. Assaraf, M. Caffarel, and A. Khelif. Diffusion Monte Carlo methods with a fixed number of walkers. *Phys. Rev. E*, vol. 61, pp. 4566–4575 (2000).

[19] S. Asmussen, and H. Hering. *Branching Processes*. Birkhäuser, Boston (1983).

[20] S. Asmussen, P.W. Glynn. *Stochastic Simulation. Algorithms and Analysis*. Springer Series: Stochastic modelling and applied probability, vol. 57 (2007).

[21] K. B. Athreya and P.E. Ney. *Branching Processes*, Springer-Verlag, NewYork (1972).

[22] S. K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277 (2001).

[23] S. K. Au and J. L. Beck. Subset simulation and its application to seismic risk based on dynamic analysis. *Journal of Engineering Mechanics ASCE*, pp. 901–917 (2003).

[24] S. K. Au, J. Ching, J. L. Beck. Application of subset simulation methods to reliability benchmark problems. *Structural Safety*, vol. 29, no. 3, pp.183–193 (2007).

[25] R. J. Aumann Markets with a continuum of traders. *Econometrica*, vol. 32, No. 1-2, pp. 39–50 (1964).

[26] A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan. *Case Studies in Spatial Point Process Modeling*, Springer Lecture Notes in Statistics 185, Springer-Verlag, New York (2006).

[27] C. Baehr, C. Beigbeder, F. Couvreux, A. Dabas, and B. Piguet. Retrieval of the turbulent and backscattering properties atmospheric properties using a nonlinear filtering technique applied to Doppler Lidar observations. Submitted to *Journal of Atmospheric and Oceanic Technology* (2012).

[28] C. Baehr. Nonlinear filtering for observations on a random vector field along a random path. M2AN - ESAIM, vol. 44, no. 5, pp. 921–945 (2010).

[29] C. Baehr. Probabilistic models for atmospheric turbulence to filter measurements with particle approximations, Ph.D. thesis, Paul Sabatier University, Toulouse, Speciality: Applied Mathematics, Option Probability (2008).

[30] C. Baehr and O. Pannekoucke. Some issues and results on the EnKF and particle filters for meteorological models in *Chaotic Systems: Theory and Applications.*, C. H. Skiadas and I. Dimotikalis, Eds. World Scientific 27-34. DOI No: 10.1142/9789814299725-0004 (2010).

[31] C. Baehr. Stochastic modeling and filtering of discrete measurements for a turbulent field. application to measurements of atmospheric wind. Int. Journal Modern Phys. B, vol. 23, no. 28-29, pp. 5424–5433 (2009).

[32] A. Bain, D. Crisan. *Fundamentals of Stochastic Filtering*, Springer Series: Stochastic Modelling and Applied Probability, vol. 60, Springer Verlag (2008).

[33] E. Balkovsky, G. Falkovich, and A. Fouxon. Intermittent distribution of inertial particles in turbulent flows, *Phys. Rev. Lett.*, vol. 86, 2790–2793 (2001).

[34] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. *Probability Theory and Related Fields*, vol. 104, no. 1, pp. 43–60 (1996).

[35] A. L. Barabàsi, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1-2, pp. 173–187 (1999)

[36] O. Bardou, S. Bouthemy, and G. Pagès. Optimal quantization for the pricing of swing options. *Applied Mathematical Finance*, vol. 16, Issue 2, pp. 183–217 (2009).

[37] J. Barral, R. Rhodes, V. Vargas. Limiting laws of supercritical branching random walks. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, vol. 350, no 9-10, pp. 535–538 (2012).

[38] S. Barthelmé and N. Chopin. ABC-EP: Expectation Propagation for Likelihood-free Bayesian Computation, ICML 2011 (Proceedings of the 28th International Conference on Machine Learning), L. Getoor and T. Scheffer (Eds.), 289–296 (2011).

[39] N. Bartoli and P. Del Moral. *Simulation et aux Algorithmes Stochastiques.* Cépaduès Édition (2001).

[40] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov Chains. *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563 (1966).

[41] L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, vol. 73, no. 3, pp. 360–363 (1967).

[42] L. E. Baum and G. R. Sell. Growth transformations for functions on manifolds. *Pacific Journal of Mathematics*, vol. 27, no. 2 pp. 211–227 (1968).

[43] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov Chains. *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171 (1970).

[44] L. E. Baum. An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process. *Inequalities*, vol. 3, pp. 1–8. (1972).

[45] E. R. Beadle and P. Djuric. A fast-weighted Bayesian bootstrap filter for nonlinear model state estimation. *IEEE Transactions on Aerospace and Electronic Systems.* vol. 33, no. 1, pp. 338–343 (1997).

[46] N. Bellomo, M. Pulvirenti. *Modeling in Applied Sciences: A kinetic theory approach.* Birkhaüser Boston (2000).

[47] A. Bellouquid and M. Delitala. Mathematical modeling of Complex biological systems. A kinetic theory approach. Birkhaüser, Boston (2006).

[48] C. Bender and J. Steiner. Least-squares Monte Carlo for backward SDEs. in *Numerical Methods in Finance.* (2011).

[49] S. Ben Hamida and R. Cont. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance*, Vol. 8, No. 4 (2005).

[50] A. Bensoussan and J. Frehse. Stochastic games for N players. *Journal of Optimization Theory and Applications*, vol. 105, no. 3, pp. 543–565 (2000).

[51] J. Berestycki, N. Berestycki, and J. Schweinsberg. Critical branching Brownian motion with absorption: survival probability. *ArXiv:1212.3821v2 [math.PR]* (2012)

[52] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series: Statistics (2nd ed.). Springer-Verlag (1985).

[53] J. M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley (1994).

[54] J. D. Biggins. Uniform convergence of martingales in the branching random walk. *Ann. Probab.* vol. 20, pp. 137–151 (1992).

[55] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York (1986).

[56] K. Binder. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford University Press, New York (1995).

[57] T. Björk, M. H. A. Davis, and C. Landèn. Optimal investment under partial information. *Mathematical Methods of Operations Research* (2010).

[58] S. Boi, V. Capasso, and D. Morale. Modeling the aggregative behavior of ants of the species Polyergus rufescens. *Nonlinear Analysis: Real World Applications*, vol. 1, pp. 163–176 (2000).

[59] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly self consistent Vlasov-Fokker-Planck Equation *ESAIM Mathematical Modelling and Numerical Analysis*, vol. 44, no. 5, 867–884 (2010).

[60] L. Boltzmann. *Leçons sur la Théorie des Gaz*. Gauthiers-Villars, tome I, (1902), tome II (1905), Réédition, Jean Gabay, Paris (1987).

[61] V. S. Borkar. Controlled diffusion processes. *Probability Surveys*, vol. 2, pp. 213-244 (2005).

[62] M. Born; J.R. Oppenheimer. On the Quantum Theory of Molecules. Annalen der Physik (in German), vol. 389, no. 20, pp. 457–484 (1927).

[63] L. Bornn, P. Jacob, P. Del Moral, and A. Doucet. An adaptive interacting Wang-Landau algorithm for automatic density exploration. arXiv:1109.3829v2 [stat.CO] ArXiv April. 2012 [33p] . To appear in *Journal of Computational and Graphical Statistics* (2013).

[64] M. Bossy and D. Talay. A stochastic particle method for some one-dimensional nonlinear PDE. *Mathematics and Computer in Simulation*, vol. 38, pp. 43–50 (1995).

[65] M. Bossy and D. Talay. Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *Ann. Appl.Probab.*, vol. 6, pp. 818–861 (1996).

[66] M. Bossy and D. Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation*. vol. 66, no. 217, pp. 157–192 (1997).

[67] B. Bouchard and X. Warin. Monte-Carlo valorisation of American options: facts and new algorithms to improve existing methods. *Numerical Methods in Finance*, Springer (2011).

[68] A. Bouchard-Côté, and S. Sankararaman, M. I. Jordan. Phylogenetic Inference via Sequential Monte Carlo. *Systematic Biology*, vol. 61, no. 4, pp. 579–593 (2012).

[69] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Process Appl.* 111, pp. 175–206 (2004).

[70] J. P. Bouchaud and M. Potters. Back to basics: historical option pricing revisited. *Philosophical Transactions: Mathematical, Physical & Engineering Sciences*, 357(1758), pp. 2019–2028 (2002).

[71] P. Bougerol and J. Lacroix. *Products of Random Matrices with Applications to Schrödinger Operators*. Birkhäuser, Boston (1985).

[72] N. Bou-Rabee and M. Hairer. Nonasymptotic mixing of the MALA algorithm. *IMA Journal of Numerical Analysis* (2012).

[73] R. D. Bourgin and R. Cogburn On determining absorption probabilities for Markov Chains in random environments. *Advances in Applied Probability*, vol. 13, no. 2, pp. 369–387 (1981).

[74] P. P. Boyle and D. Emanuel. Discretely adjusted option hedges. *Journal of Financial Economics*, vol. 8, pp. 259–282 (1980).

[75] R. N. Bradt, S. M. Johnson, and S. Karlin. On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.*, vol. 27, pp. 1060–1070 (1956).

[76] M. D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, vol 31, pp. 531–581 (1978).

[77] L. Breiman. *Probability*. Original edition published by Addison-Wesley, 1968; reprinted by Society for Industrial and Applied Mathematics (1992).

[78] J. Bretagnolle. A new large deviation inequality for U-statistics of order 2. *ESAIM: Probability and Statistics*. Vol. 3, pp. 151–162 (1999).

[79] D. Brigo, T. Bielecki, and Fr. Patras. *Credit Risk Frontiers*. Wiley–Bloomberg Press (2011).

[80] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC Press, col. 25, iss. 1 (2012).

[81] J. K. Brooks. Representations of weak and strong integrals in Banach spaces. *Proc. Nat. Acad. Sci. U.S.A.*, vol. 63, pp. 266–270 (1969).

[82] M. Broadie and P. Glasserman. Estimating security prices using simulation. *Management Science*, 42, pp. 269–285 (1996).

[83] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, vol. 7, pp. 35–72, (2004).

[84] F. T. Bruss. The art of a right decision: Why decision makers want to know the odds-algorithm. *Newsletter of the European Mathematical Society*, Issue 62, pp. 14–20 (2006).

[85] A. Budhiraja, P. Del Moral, and S. Rubenthaler. Discrete time Markovian agents interacting through a potential. *Journal ESAIM Probability & Statistics* (2012).

[86] R. Buckdahn, B. Djehiche, J. Li, and S. Peng. Mean-field backward stochastic differential equations: a limit approach. *The Annals of Probability*, vol. 37, no. 4, pp. 15241–565 (2009).

[87] R. Buckdahn, B. Djehiche, J. Li, and S. Peng. Mean-field backward stochastic differential equations and related partial differential equations. *Stochastic Processes and Their Applications*, vol. 119, no. 10, pp. 3133–3154 (2007).

[88] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. John Wiley & Sons, New York (1990).

[89] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer-Verlag (2004).

[90] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer-Verlag, New York (2004).

[91] G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724 (1998).

[92] M. Caffarel, R. Assaraf, A. Khelif, A. Scemama, A. Ramirez-Solis. A few aspects of QMC for molecules. In Mathematical and Numerical Aspects of Quantum Chemistry Problems, Mathematisches Forschunginstitut Oberwolfach, p.7 Report No.47/2006 (2006).

[93] M. Caffarel, R. Assaraf. A pedagogical introduction to quantum Monte Carlo. In Mathematical models and methods for ab initio Quantum Chemistry in Lecture Notes in Chemistry, eds. M. Defranceschi and C.Le Bris, Springer p.45 (2000).

[94] M. Caffarel. Stochastic methods in quantum mechanics. In Numerical Determination of the Electronic Structure of Atoms, Diatomic and Polyatomic Molecules, pp.85-105 Kluwer Academic Publishers, (1989).

[95] F. Campillo and V. Rossi. Convolution particle filtering for parameter estimation in general state-space models. In *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego (USA), December (2006).

[96] F. Campillo and V. Rossi. Convolution filter based methods for parameter estimation in general state space models. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1063–1071 (2009).

[97] E. Cancès, B. Jourdain, and T. Lelièvre. Quantum Monte Carlo simulations of fermions. A mathematical analysis of the fixed-node approximation, *ESAIM: M2AN*, vol. 16, no. 9, pp. 1403–1449, (2006).

[98] E. Cancès, F. Legoll, and G. Stoltz. Theoretical and numerical comparison of sampling methods for molecular dynamics. *ESAIM: M2AN*, vol. 41, no. 2, pp. 351–389 (2005).

[99] O. Cappé, E. Moulines, and T. Rydèn. *Inference in Hidden Markov Models*, Springer-Verlag (2005).

[100] O. Cappe, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in SMC. *Proceedings of the IEEE*, vol. 95, no. 5, pp.899–924 (2007)

[101] R. Carmona, and S. Crépey. Particle methods for the estimation of credit portfolio loss distributions. *International Journal of Theoretical and Applied Finance*, vol.13, no. 4, pp. 577–602 (2010).

[102] R. Carmona, P. Del Moral, P. Hu, and N. Oudjane. An introduction to particle methods in finance. In *Numerical Methods in Finance*. Numerical Methods in Finance Series: Springer Proceedings in Mathematics, vol. 12, no. XVII, p. 3–49 (2012).

[103] R. Carmona, P. Del Moral, P. Hu, and N. Oudjane. Numerical Methods in Finance Series: Springer Proceedings in Mathematics, Bordeaux, June 2010 Vol. 12, no. XVII (2012).

[104] R. Carmona, J.-P. Fouque, and D. Vestal. Interacting Particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, vol. 13, no. 4, pp. 613–633 (2009).

[105] R. Carmona, J.-P. Fouque, and L.-H. Sun: Mean Field Games and Systemic Risk, submitted 2013, revised 2014, to appear in Communications in Mathematical Sciences.

[106] R. Carmona and N. Touzi. Optimal multiple stopping and valuation of swing options. *Mathematical Finance*, vol. 18, no. 2, pp. 239–268 (2008).

[107] F. Caron, P. Del Moral, A. Tantar, and E. Tantar. Simulation particulaire, Dimensionnement en conditions extrêmes. *Research contract no. 2010-IFREMER-01* with IFREMER, (92p) September (2010).

[108] F. Caron, P. Del Moral, A. Doucet, and M. Pace. On the conditional distributions of spatial point processes. *Adv. in Appl. Probab.*, vol. 43, no. 2, pp. 301–307. (2011).

[109] F. Caron, P. Del Moral, M. Pace, and B. N. Vo. On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering stoch. *Analysis and Applications* vol. 29, no. 6, p. 951–997 (2011).

[110] F. Caron, P. Del Moral, A. Doucet, and M. Pace. Particle approximations of a class of branching distribution flows arising in multi-target tracking *SIAM J. Control Optim.*, vol. 49, pp. 1766–1792 (2011).

[111] J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for nonlinear problems. *IEE Proceedings* F, vol. 146, pp. 2–7 (1999).

[112] J. F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, vol. 19, pp. 19–30 (1996).

[113] R. Casarin. Simulation methods for nonlinear and non-Gaussian models in finance. *Premio SIE, Rivista Italiana degli Economisti*, vol. 2, pp. 341–345 (2005).

[114] R. Casarin and C. Trecroci. Business cycle and stock market volatility: a particle filter approach. *Cahier du CEREMADE N. 0610*, University Paris Dauphine (2006).

[115] B. Cascales, M. Raja. Measurable selectors for the metric projection. Mathematische Nachrichten, pp. 254–255, vol. 1 (2003).

[116] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. S. Martin. Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.*, vol. 37, no. 5, pp. 1926–1969 (2009).

[117] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the nonuniformly convex case. *Probability Theory and Related Fields.* vol. 140, no.1–2, pp. 19–40 (2008).

[118] J. A. Cavender. Quasi-stationary distributions of birth-and-death processes. *Adv. Appl. Probab.*, vol. 10, no. 3, pp. 570–586 (1978).

[119] C. Cercignani. *The Boltzmann Equation and Its Applications.* Springer-Verlag, New York (1988).

[120] C. Cercignani. *Rarefied Gas Dynamics.* Cambridge University Press, Cambridge (2000).

[121] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases.* Springer (1994).

[122] A. Cerny. Dynamic programming and mean variance hedging in discrete time. *Tanaka Business School Discussion Papers* TBS/DP04/15 London: Tanaka Business School (2004).

[123] R. Cerf. Une théorie asymptotique des algorithmes génétiques, Ph.D., Université Montpellier II (1994).

[124] R. Cerf. Asymptotic convergence of a genetic algorithm, *C.R. Acad. Sci. Paris*, no. 319, Série I, pp. 271–276 (1994).

[125] R. Cerf. Asymptotic convergence of genetic algorithms. *Adv. Appl. Probab.*, vol. 30, pp. 521–550 (1998).

[126] R. Cerf. A new genetic algorithm, *Annals of Applied Probab.* vol. 3, pp. 778–817 (1996).

[127] R. Cerf. The dynamics of mutation-selection algorithms with large population sizes, *Ann. Inst. Henri Poincaré*, vol. 32 no.4, pp. 455–508 (1996).

[128] R. Cerf. An asymptotic theory for genetic algorithms, *Artificial Evolution*, Lecture Notes in Computer Science vol. 1063, pp. 37–53, Springer-Verlag (1996).

[129] F. Cerou, A. Guyader, R. Rubinstein, and R. Vaismana. Smoothed splitting method for counting. *Stochastic models* (2011), vol. 10, no. 27, pp. 626–650.

[130] F. Cérou, P. Del Moral, and A. Guyader. A nonasymptotic variance theorem for unnormalized Feynman-Kac particle models. Annales Institute Henri Poincoré, Proba. and Stat., vol. 47, no. 3, pp 629–669 (2011).

[131] F. Cerou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, vol. 22, no. 3, pp. 795–808 (2012).

[132] F. Cérou, P. Del Moral, F. LeGland, and P. Lézaud. Limit Theorems for multilevel splitting algorithms in the simulation of rare events. *Proceedings of the 2005 Winter Simulation Conference.* M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, Eds. (2005).

[133] F. Cérou, P. Del Moral, F. Le Gland, and P. Lézaud. Genealogical models in entrance times rare event analysis. *Alea.* vol. 1, pp. 181–203 (2006).

[134] P. M. Chaikin, T. C. Lubensky. *Principles of Condensed Matter Physics.* Cambridge University Press, London (2007).

[135] M. Chaleyat-Maurel and D. Michel. Des résultats de nonexistence de filtres de dimension finie. *C. R. Acad. Sc. de Paris Série I Math.*, no. 296, no. 22, pp. 933–936 (1983).

[136] S. Chapman and T. Cowling. *The Mathematical Theory of Non-Uniform Gases.* Cambridge University Press (1952).

[137] R. Chelli, S. Marsili, A. Barduci, and P. Procacci. Generalization of the Jarzynski and Crooks nonequilibrium work theorems in molecular dynamics simulations *Phys. Rev. E*, vol. 75, no. 5, 050101(R) (2007)

[138] L. Y. Chen. On the Crooks fluctuation theorem and the Jarzynski equality. *J Chem Phys.*, vol. 129, no. 9, 091101 (2008)

[139] N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and Their Applications*, vol. 117, pp. 1689–1723 (2007).

[140] R. Chen and J. Liu. Mixture Kalman filters. *J. Royal Statistical Society B*, vol. 62, pp. 493–508 (2000).

[141] N. Chopin. A sequential particle filter method for static models. *Biometrika*, vol. 89, pp. 539–552 (2002).

[142] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2: An efficient algorithm for sequential analysis of state-space models. arXiv:1101.1528v3 (2012). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2013).

[143] N. Chopin, T. Lelièvre, and G. Stoltz. Free energy methods for efficient exploration of mixture posterior densities. *Statistics and Computing*, vol. 22, no. 4, pp. 897–916 (2012).

[144] Y. S. Chow, H. Robbins and D. Siegmund. *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston (1971).

[145] D. E. Clark and J. Bell. Convergence results for the particle PHD filter, *Signal Processing, IEEE Transactions*, volume 54, issue 7, pp. 2652–2661, (2006).

[146] E. Cleland, G. H. Booth, and A. Alavi. Communications: survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo. *The Journal of Chemical Physics*, vol. 132, 041103 (2010).

[147] E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, vol. 6, pp. 449–472, (2002).

[148] J. Cohen. Products of random matrices and related topics in mathematics and science: a bibliography. *Contemp. Math.*, vol. 50, pp. 337–358 (1986).

[149] F. Comets, T. Shiga, and N. Yoshida. Probabilistic analysis of directed polymers in a random environment: A review. *Adv. Stud. in Pure Math.* (2004).

[150] B. Cooke, J.C. Mattingly, S.A. McKinley, S.C. Schmidler. Geometric ergodicity of the two dimensional Hamiltonian systems with a Lennard Jones like repulsive potential. eprint arXiv:1104.3842 (2011).

[151] D. D. Creal. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, vol. 31, no. 3, pp. 245–296. (2012).

[152] D. Crisan and B. Rozovsky (Eds). *Handbook of Nonlinear Filtering*, Cambridge University Press (2009).

[153] D. Crisan, P. Del Moral, and T. Lyons. Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields*, vol. 5, no. 3, pp. 293–318 (1999).

[154] D. Crisan, P. Del Moral, and T. Lyons. Interacting particle systems approximations of the Kushner Stratonovitch equation. *Advances in Applied Probability*, vol. 31, no. 3, pp. 819–838 (1999).

[155] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, vol. 90, 1481 (1998).

[156] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, Springer, New York (1988).

[157] D. A. Dawson. Critical dynamics and fluctuations for a mean field model of cooperative behavior. *J. Statistical Physics*, vol. 31, no. 1, pp. 29–85 (1983).

[158] D. A. Dawson, and P. Del Moral. Large deviations for interacting processes in the strong topology. *Statistical Modeling and Analysis for Complex Data Problem*. P. Duchesne and B. Rémillard, Eds. Springer, pp. 179–209 (2005).

[159] M. deBruijne, and M. Nielsen. Shape particle filtering for image segmentation. In Proceedings, Medical Image Computing and Computer-Assisted Intervention, pp.168–175 (2004).

[160] P. Del Moral. Nonlinear filtering: interacting particle solution. *Markov Processes and Related Fields*, vol. 2, no. 4, pp. 555–580 (1996).

[161] P. Del Moral. Measure valued processes and interacting particle systems. Application to nonlinear filtering problems. *Annals of Applied Probability*, vol. 8, no. 2, pp. 438–495 (1998).

[162] P. Del Moral. Fundamentals of stochastic filtering by Alan Bain and Dan Crisan Book Review [13p.], *Bulletin of the American Mathematical Society*, October (2010).

[163] P. Del Moral, J. Jacod, and P. Protter. The Monte-Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields*, vol. 120, pp. 346–368 (2001).

[164] P. Del Moral and J. Jacod. The Monte-Carlo Method for filtering with discrete time observations. Central Limit Theorems. *The Fields Institute Communications, Numerical Methods and Stochastics*, T.J. Lyons T.S. Salisbury, and Eds., American Mathematical Society, (2002).

[165] P. Del Moral and J. Jacod. Interacting particle filtering with discrete observations. In *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science. Springer. A. Doucet, J. F. G. de Freitas, N. J. Gordon., Eds., pp. 43–77 (2001).

[166] P. Del Moral, L. Kallel, and J. Rowe. Modeling genetic algorithms with interacting particle systems. *Revista de Matematica, Teoria y aplicaciones*, vol. 8, no. 2, July (2001).

[167] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling procedures for sequential Monte Carlo methods (HAL-INRIA RR-6700) [46p], (Oct. 2008). *Bernoulli*, vol. 18, no. 1, pp. 252–278 (2012).

[168] P. Del Moral, A. Doucet, and A. Jasra. An adaptive sequential Monte Carlo method for approximate Bayesian computation (DOI: 10.1007/s11222-011-9271-y). *Statistics and Computing*, vol. 1–12, August 03 (2011).

[169] P. Del Moral and M. Ledoux. On the convergence and the applications of empirical processes for interacting particle systems and nonlinear filtering. *Journal of Theoretical Probability*, vol. 13, no. 1, pp. 225–257 (2000).

[170] P. Del Moral and L. Miclo. *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering.* Sminaire de Probabilits XXXIV, J. Azma and M. Emery and M. Ledoux and M. Yor, Lecture Notes in Mathematics, Springer-Verlag, Berlin, vol. 1729, pp. 1–145 (2000).

[171] P. Del Moral and L. Miclo. On the convergence and the applications of the generalized simulated annealing. *SIAM Journal on Control and Optimization*, vol. 37, no. 4, pp. 1222–1250, (1999).

[172] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer-Verlag, New York (2004).

[173] P. Del Moral and A. Doucet. *On a Class of Genealogical and Interacting Metropolis Models.* Séminaire de Probabilités XXXVII, J. Azéma and M. Emery and M. Ledoux and M. Yor, Eds., Lecture Notes in Mathematics 1832, Springer-Verlag, Berlin, pp. 415–446 (2003).

[174] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *J. Royal Statist. Soc.* B, vol. 68, pp. 411–436 (2006).

[175] P. Del Moral and E. Rio. Concentration inequalities for mean field particle models. *Ann. Appl. Probab*, vol. 21, no. 3, pp. 1017–1052. (2011).

[176] P. Del Moral and L. Miclo. Annealed Feynman-Kac models. *Comm. Math. Phys.*, vol. 235, pp. 191–214 (2003).

[177] P. Del Moral, and A. Guionnet. Large deviations for interacting particle systems. Applications to nonlinear filtering problems. *Stochastic Processes and Their Applications*, vol. 78, pp. 69–95 (1998).

[178] P. Del Moral and A. Guionnet. On the stability of measure valued processes with applications to filtering. *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 329, pp. 429–434 (1999).

[179] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré*, vol. 37, no. 2, pp. 155–194 (2001).

[180] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, vol. 15, no. 4, pp. 2496–2534 (2005).

[181] P. Del Moral and J. Garnier. Simulations of rare events in fiber optics by interacting particle systems. *Optics Communications* vol. 257, pp. 205–214 (2006)

[182] P. Del Moral and N. Hadjiconstantinou. An Introduction to probabilistic methods, with applications. vol. 44, no. 5, pp. 805–830 M2AN (2010).

[183] P. Del Moral, S. Hu, and L.M. Wu. Moderate deviations for mean field particle models HAL-INRIA no. 00687827 [37p] (2012) .

[184] P. Del Moral and A. Doucet. Particle motions in absorbing medium with hard and soft obstacles. *Stochastic Anal. Appl.*, vol. 22, pp. 1175–1207 (2004).

[185] P. Del Moral and L. Miclo. Particle approximations of Lyapunov exponents connected to Schrdinger operators and Feynman-Kac semigroups. *ESAIM: Probability and Statistics*, no. 7, pp. 171–208 (2003).

[186] P. Del Moral and L. Miclo. Genealogies and Increasing Propagations of Chaos for Feynman-Kac and Genetic Models. *Annals of Applied Probability*, vol. 11, no. 4, pp. 1166–1198 (2001).

[187] P. Del Moral and L. Miclo. Strong propagations of chaos in Moran's type particle interpretations of Feynman-Kac measures. *Stochastic Analysis and Applications*, vol. 25, no. 3, pp. 519–575 (2007).

[188] P. Del Moral and L. Miclo. A Moran particle system approximation of Feynman-Kac formulae. *Stochastic Processes and Their Applications*, vol. 86, pp. 193–216 (2000).

[189] P. Del Moral and L. Miclo. On the Stability of Non Linear Semigroup of Feynman-Kac Type. *Annales de la Faculté des Sciences de Toulouse*, vol. 11, no. 2, pp. 135–175 (2002)

[190] P. Del Moral and L. Miclo. Asymptotic Results for Genetic Algorithms with Applications to Non Linear Estimation. *Proceedings Second EvoNet Summer School on Theoretical Aspects of Evolutionary Computing*. Ed. B. Naudts, L. Kallel. Natural Computing Series, Springer (2000).

[191] P. Del Moral, M. Ledoux, and L. Miclo. Contraction properties of Markov kernels. *Probab. Theory and Related Fields*, vol. 126, pp. 395–420, (2003).

[192] P. Del Moral, L. Miclo, F. Patras, and S. Rubenthaler. The convergence to equilibrium of neutral genetic models. *Stochastic Analysis and Applications*. vol. 28, no. 1, pp. 123–143 (2009).

[193] P. Del Moral and P. Lezaud. Branching and interacting particle interpretation of rare event probabilities. *Stochastic Hybrid Systems: Theory and Safety Critical Applications*, H. Blom and J. Lygeros, Eds. Springer-Verlag, Heidelberg (2006).

[194] P. Del Moral, P. Jacob, A. Lee, L. Murray, and G. Peters. Feynman-Kac particle integration with geometric interacting jumps. arXiv:1211.7191 November (2012).

[195] P. Del Moral, P. Hu, and L.M. Wu. On the concentration properties of interacting particle processes. *Foundations and Trends in Machine Learning*, vol. 3, no. 3-4, pp. 225–389 (2012).

[196] P. Del Moral and F. Patras. Interacting path systems for credit risk. *Credit Risk Frontiers*. D. Brigo, T. Bielecki, and F. Patras, Eds. *Wiley–Bloomberg Press*, (2011), 649–674. *Short announcement available as Interacting path systems for credit portfolios risk analysis. INRIA:RR-7196 (2010).*

[197] P. Del Moral, F. Patras., and R. Kohn On Feynman-Kac and particle Markov chain Monte Carlo models. arXiv preprint arXiv:1404.5733 (2014).

[198] P. Del Moral, F. Patras., and S. Rubenthaler. Convergence of U-statistics for interacting particle systems. *Journal of Theoretical Probability*, vol. 24, no. 4, pp 1002–1027 december (2011).

[199] P. Del Moral, Fr. Patras., and S. Rubenthaler. Coalescent tree based functional representations for some Feynman-Kac particle models. *Annals of Applied Probability*, vol. 19, no. 2, pp. 1–50 (2009).

[200] P. Del Moral, Fr. Patras., and S. Rubenthaler. A mean field theory of nonlinear filtering. *Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Cambridge University Press (2009).

[201] P. Del Moral, Br. Rémillard, and S. Rubenthaler. Monte Carlo approximations of American options that preserve monotonicity and convexity. In *Numerical Methods in Finance*. Numerical Methods in Finance Series: Springer Proceedings in Mathematics, vol. 12, no. XVII (2012).

[202] P. Del Moral and A. Doucet. Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations. *The Annals of Applied Probability*, vol. 20, no. 2, pp. 593–639 (2010)

[203] P. Del Moral, A. Doucet, and S. S. Singh. A Backward particle interpretation of Feynman-Kac formulae *M2AN ESAIM*. vol 44, no. 5, pp. 947–976 (Sept. 2010)

[204] P. Del Moral, A. Doucet, and S. S. Singh. Forward smoothing using sequential Monte Carlo. *Technical Report CUED/F-INFENG/TR 638. Cambridge University Engineering Department* (2009).

[205] P. Del Moral, A. Doucet, and S. S. Singh. Computing the filter derivative using Sequential Monte Carlo. *Technical Report. Cambridge University Engineering Department* (2011).

[206] P. Del Moral, P. Hu, and N. Oudjane. Snell envelope with small probability criteria. *Applied Mathematics and Optimization*, vol. 66, no. 3, pp 309–330, December (2012).

[207] P. Del Moral, P. Hu, N. Oudjane, Br. Rémillard. On the robustness of the Snell envelope. *SIAM Journal on Financial Mathematics*, vol.2, pp. 587–626 (2011).

[208] P. Del Moral, M.A. Kouritzin, and L. Miclo. On a class of discrete generation interacting particle systems. *Electronic Journal of Probability*, no. 16, no. 26 (2001).

[209] P. Del Moral, and S. Tindel. A Berry-Esseen theorem for Feynman-Kac and interacting particle models. *Annals of Applied Probability*, vol. 15, no. 1B, pp. 941–962 (2005).

[210] P. Del Moral, J. Tugaut. Uniform propagation of chaos for a class of inhomogeneous diffusions. *hal-00798813* (2013).

[211] P. Del Moral, Ch. Vergé. Stochastic models and methods. An introduction with applications (in french) Springer Series : Maths & Applications, SMAI, vol. 75, vol. 75 (2014).

[212] B. E. Brunet, B. Derrida, A. H. Muller, and S. Munier. Effect of selection on ancestry: and exactly soluble case and its phenomenological generalization. *Phys. Review E*, vol. 76, 041104 (2007).

[213] B. Derrida, M. R. Evans, and E. R. Speer. Mean field theory of directed polymers with random complex weights. *Commun. Math. Phys.*, vol. 156, pp. 221–244 (1993).

[214] B. Derrida, and D. Simon. The survival probability of a branching random walk in the presence of an absorbing wall. *Europhys. Lett. EPL*, vol. 78, Art. 60006 (2007).

[215] B. Derrida, and D. Simon. Quasi-stationary regime of a branching random walk in presence of an absorbing wall. *J. Statist. Phys.*, vol. 131, pp. 203–233 (2008).

[216] B. Derrida, and H. Spohn. Polymers on disordered trees, spin glasses, and traveling waves. *Journal of Statistical Physics*, vol. 51, nos. 5/6, pp. 817–840 (1988).

[217] P. Diaconis. Some things we've learned (about Markov chain Monte Carlo). Preprint Stanford University (2012).

[218] P. Diaconis, G. Lebeau, and L. Michel. Gibbs/Metropolis algorithm on a convex polytope. *Math. Zeitschrift*, published online: Aug. (2011).

[219] P. Diaconis, G. Lebeau, and L. Michel. Geometric Analysis for the Metropolis Algorithm on Lipschitz Domains. *Invent. Math.*, vol. 185, no. 2, pp. 239–281 (2010).

[220] P. Diaconis and L. Miclo. On characterizations of Metropolis type algorithms in continuous time. *Alea*, vol. 6, pp. 199–238 (2009).

[221] P. Diaconis and G. Lebeau. Micro-local analysis for the Metropolis algorithm. *Mathematische Zeitschrift*, vol. 262, no. 2, pp. 441–447 (2009).

[222] P. Diaconis. The Markov Chain Monte Carlo revolution. *Bull. Amer. Math. Soc.*, November (2008).

[223] P. Diaconis and F. Bassetti. Examples comparing importance sampling and the Metropolis algorithm. *Illinois Journal of Mathematics*, vol. 50, no. 1–4, pp. 67–91 (2006).

[224] P. Diaconis J. Neuberger. Numerical results for the Metropolis algorithm. *Experimental Math.*, vol. 13, no. 2, pp. 207–214 (2004).

[225] P. Diaconis and L. Billera. A geometric interpretation of the Metropolis-Hastings algorithm. *Statis. Sci.*, vol. 16, no. 4, pp. 335–339 (2001).

[226] P. Diaconis and A. Ram. Analysis of systematic scan Metropolis Algorithms using Iwahori-Hecke algebra techniques. *Michigan Journal of Mathematics*, vol. 48, no. 1, pp. 157–190 (2000).

[227] P. Diaconis, L. Saloff-Coste. What do we know about the Metropolis algorithm? *Jour. Comp. System Sciences*, vol. 57, pp. 20–36 (1998).

[228] P. Diaconis and P. Hanlon. Eigen-analysis for some examples of the Metropolis algorithm. *Contemporary Math.*, vol. 138, pp. 99–117 (1992).

[229] P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*, 2nd ed., Arnold (2003).

[230] P. Dirac. A new notation for quantum mechanics. Mathematical Proceedings of the Cambridge Philosophical Society, vol. 35, no. 3, pp. 416–418 (1939).

[231] R. L. Dobrushin. Description of a random field by means of conditional probabilities, and the conditions governing its regularity. *Theor. Proba. Appli.*, vol. 13, pp. 197–224 (1968).

[232] R. Douc, A. Garivier, E. Moulines, and J. Olsson. On the forward filtering backward smoothing particle approximations of the smoothing distribution in general state spaces models. *Annals of Applied Probab.*, vol. 21, no. 6, pp. 2109–2145 (2011).

[233] A. Doucet, B. Vo, and S. S. Singh. Sequential Monte Carlo methods for Bayesian multi-target filtering with random finite sets. *IEEE Trans. Aerospace Elec. Systems*, vol. 41, pp. 1224–1245 (2005).

[234] A. Doucet, J.F.G. de Freitas, and N.J. Gordon Eds. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York (2001).

[235] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later in *Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds., Cambridge University Press (2009).

[236] A. Doucet, L. Montesano, and A. Jasra. Optimal filtering for partially observed point processes, *Proceedings ICASSP* (2006).

[237] A. Doucet, S.J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, vol. 10, pp. 197–208 (2000).

[238] C. Dubarry and S. Le Corff. Nonasymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models, Preprint (2012).

[239] C. Dubarry, P. Del Moral, and E. Moulines, Ch. Vergé. On the convergence of Island particle models, Preprint (2012).

[240] P. Dupuis and H. Wang. Importance sampling for Jackson networks. *Queueing Systems*, vol. 62, pp. 113–157 (2009).

[241] P. Dupuis, K. Leder, and H. Wang. Importance sampling for weighted-serve-the-longest-queue, *Math. of Operations Research*, vol. 34, pp. 642–660 (2009).

[242] P. Dupuis, K. Leder, and H. Wang. Large deviations and importance sampling for a tandem network with slow-down. *QUESTA*, vol. 57, pp. 71–83 (2007).

[243] P. Dupuis, K. Leder, and H. Wang. Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. on Modelling and Computer Simulation*, vol. 17, pp. 1–21 (2007).

[244] P. Dupuis, D. Sezer, and H. Wang. Dynamic importance sampling for queueing networks. *Annals of Applied Probability*, vol. 17, pp. 1306–1346 (2007).

[245] P. Dupuis and H. Wang. Sub-solutions of an Isaacs equation and efficient schemes for importance sampling. *Mathematics of Operations Research*, vol. 32, pp. 1–35 (2007).

[246] P. Dupuis and H. Wang. Optimal stopping with random intervention times. *Adv. Applied Probability*, vol. 34, pp.1–17 (2002).

[247] P. Dupuis and H. Wang. On the convergence from discrete to continuous time in an optimal stopping problem. *Annals of Applied Probability*, vol. 15, pp. 1339–1366 (2005).

[248] P. Dybvig. Distributional analysis of portfolio choice. *The Journal of Business*, vol. 61, no. 3, pp. 369–393 (1988).

[249] A. Eberle and C. Marinelli. Quantitative approximations of evolving probability measures and sequential MCMC methods. *Probability Theory and Related Fields*, pp. 1–37 (2008).

[250] A. Eberle and C. Marinelli. Stability of nonlinear flows of probability measures related to sequential MCMC methods. *ESAIM: Proceedings*, vol. 19, pp. 22–31 (2007).

[251] A. Eberle and C. Marinelli. Lp estimates for Feynman-Kac propagators with time-dependent reference measures. *Journal of mathematical analysis and applications*, vol. 365, no. 1, pp. 120–134. (2010).

[252] A. Economou. Generalized product-form stationary distributions for Markov chains in random environments with queueing applications. *Adv. in Appl. Probab.*, vol. 37, no. 1, pp. 185–211 (2005).

[253] D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Annals of Applied Probability*, 15, pp. 1–37 (2005).

[254] D. Egloff, M. Kohler, and N. Todorovic. A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *Ann. Appl. Probab.*, 17, pp. 1138–1171 (2007).

[255] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, vol. 7, no. 1, pp. 1–71 (1997).

[256] M. Ellouze, J.P. Gauchi, and J.C. Augustin. Global sensitivity analysis applied to a contamination assessment model of Listeria monocytogenes in cold smoked salmon at consumption. *Risk Anal.*, vol. 30, pp. 841–852 (2010)

[257] M. Ellouze, J.P. Gauchi, and J.C. Augustin. Use of global sensitivity analysis in quantitative microbial risk assessment: application to the evaluation of a biological time temperature integrator as a quality and safety indicator for cold smoked salmon. *Food Microbiol.*, vol. 28, no. 4, pp 755–769 (2011).

[258] M. El Makrini, B. Jourdain and T. Lelièvre. Diffusion Monte Carlo method: Numerical analysis in a simple case. *ESAIM: M2AN*, vol. 41, no. 2, pp. 189–213, (2007).

[259] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research (Oceans)*, vol. 99, no. C5, pp. 10143–10162 (1994).

[260] G. Evensen. Ensemble Kalman filter: Theoretical formulation and practical implementations. *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367 (2003).

[261] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, Berlin (2006).

[262] G. Falkovich, K. Gawedzki, and M. Vergassola, Particles and fields in fluid turbulence. *Rev. Modern Phys.*, vol. 73, pp. 913–975 (2001).

[263] Y. E. Famao, S.U. Lin, L. Shukai, and L. Shengyang. Extraction of complex object contour by particle filtering. In *IEEE Int. Geoscience and Remote Sensing Symposium*, pp. 3711–3713 (2005).

[264] P. Fearnhead. Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statistics and Computing*, vol. 18, pp. 151–171 (2008).

[265] P. Fearnhead, P. Clifford. On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society.* Series B, vol. 65, no. 4, pp. 887–899 (2003).

[266] E. Fermi and R.D. Richtmyer ote on census-taking in Monte Carlo calculations Declassified report by Enrico Fermi. From the Los Alamos Archive (1948).

[267] P. A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.*, vol. 23, no. 2, pp. 501–521 (1995).

[268] P. A. Ferrari and N. Maric. Quasi stationary distributions and Fleming-Viot processes in countable spaces. *Electron. J. Probab.,* vol. 12, no. 24, pp. 684–702 (2007).

[269] G. Fishman. *Monte Carlo. Concepts, Algorithms and Applications.* Springer-Verlag (1996).

[270] J. J. Florentin. Optimal observability and optimal control. *J. of Electronics and Control.* vol. 13, pp. 263–279 (1962).

[271] I. Florescu and F. Viens. Stochastic volatility: option pricing using a multinomial recombining tree. *Applied Mathematical Finance*, vol. 15, no. 2, pp.151–181 (2008).

[272] J. P. Fouque, G. Papanicolaou, and R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility.* Cambridge University Press (2000).

[273] J.-P. Fouque and L.-H. Sun. Systemic Risk Illustrated, Handbook on Systemic Risk, Eds J.-P. Fouque and J. Langsam. Cambridge University Press (2013).

[274] E. Fournié, J. M. Lasry, J. Lebuchoux, P. L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance, *Finance and Stochastics*, vol. 3, pp. 391–412 (1999).

[275] E. Fournié, J. M. Lasry, J. Lebuchoux, and P. L. Lions. Applications of Malliavin calculus to Monte Carlo methods in finance. II, *Finance and Stochastics*, vol. 5, pp. 201–236 (2001).

[276] H. Frauenkron, U. Bastolla, E. Gerstner, P. Grassberger, and W. Nadler. New Monte Carlo algorithm for protein folding. *Physical Review Letters*, vol. 80, no. 14, pp. 3149–3152 (1998).

[277] H. Frauenkron, M.S. Causo, and P. Grassberger. Two-dimensional self-avoiding walks on a cylinder. *Phys. Rev., E*, vol. 59, pp. 16–19 (1999).

[278] M. C. Fu, D. B. Madan, and T. Wang. Pricing continuous Asian options: A comparison of Monte Carlo and Laplace transform inversion methods. *Journal of Computational Finance*, vol. 2, pp. 49–74, (1998).

[279] J. Gärtner. On the McKean-Vlasov limit for interacting particles. *Math. Nachr.*, vol. 137, pp. 197–248 (1988).

[280] D. Gasbarra. Particle filters for counting process observations. http://www.rni.helsinki.fi/∼dag/newpart2.ps, Research report Helsinki University, (2001).

[281] J. P. Gauchi, J. P. Vila, and L. Coroller. New prediction confidence intervals and bands in the nonlinear regression model: Application to the predictive modelling in food. *Communications in Statistics, Simulation and Computation*, vol. 39, no. 2, pp. 322–330 (2009).

[282] J. P. Gauchi, C. Bidot, J.C. Augustin, and J. P. Vila. Identification of complex microbiological dynamic system by nonlinear filtering. *6th Int. Conference on Predictive Modelling in Foods, Washington DC* (2009).

[283] H. Gea and S. Asgarpoorb. Parallel Monte Carlo simulation for reliability and cost evaluation of equipment and systems. *Electric Power Systems Research*, vol. 81, no. 2, pp. 347–356 (2011).

[284] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741 (1984).

[285] V. Genon-Catalot, T. Jeantheau, and C. Laredo. Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scandinavian Journal of Statistics*, vol. 30, pp. 297–316 (2003).

[286] J. Geweke and G. Durham. Massively Parallel Sequential Monte Carlo for Bayesian Inference. Available at SSRN: http://ssrn.com/abstract=1964731 or http://dx.doi.org/10.2139/ssrn.1964731 (2011).

[287] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Yale University Press, New Haven, CT (1902)

[288] W. R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, London (1996).

[289] W. R. Gilks and C Berzuini. Following a moving target Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B*, vol. 63, no. 1, pp. 127–146 (2002).

[290] W. R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall (1996).

[291] E. Giné. *Lectures on some aspects of the bootstrap*. Lecture Notes in Maths, vol. 1665, Springer (1997).

[292] E. Giné, R. Latala, J. Zinn. Exponential and moment inequalities for U-statistics. *High Dimensional Proba. II*. Birkhäuser Boston, pp.13–38 (2000).

[293] F. Giraud, P. Del Moral. Nonasymptotic analysis of adaptive and annealed Feynman-Kac particle models. arXiv:1209.5654 (2012).

[294] F. Giraud, P. Minvielle, M. Sancandi, and P. Del Moral. Rao-Blackwellised interacting Markov chain Monte Carlo for electromagnetic scattering inversion. *Journal of Physics: Conference Series* 386 (1), 012008 (2012).

[295] M. Girolami, B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, vol. 73, no. 2, pp. 123–214 (2011).

[296] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, vol. 47, pp. 585–600 (1999).

[297] P. Glasserman and B. Yu. Number of paths versus number of basis functions in American option pricing. *Ann. Appl. Probab.*, 14, pp. 1–30 (2004).

[298] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag (2004).

[299] E. Gobet, J. P. Lemor, and X. Warin. A regression-based Monte-Carlo method for backward stochastic differential equations. *Annals Applied Probability*, vol. 15, pp. 2172–2202 (2005).

[300] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA (1989).

[301] D. A. Gomes, J. Mohr, and R. Rigao Souza. Discrete time, finite state space mean field games. *Journal de Mathématiques Pures et Appliquées*, vol. 93, issue 3, pp. 308-328 (2010).

[302] N. J. Gordon, D. Salmond, and A. F. M. Smith. A novel approach to state estimation to nonlinear non-Gaussian state estimation. *IEE Proceedings* F, vol. 40, pp. 107–113 (1993).

[303] N. J. Gordon, D. Salmond, and E. Craig. Bayesian state estimation for tracking and guidance using the bootstrap filter. *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 6, pp. 1434–1443 (1995).

[304] N. J. Gordon, A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no.1, pp. 353 –358 (1997).

[305] F. Gosselin. Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *Ann. Appl. Probab.*, vol. 11, pp. 261–284 (2001).

[306] J. Goutsias, R. Mahler, and H. Nguyen, Eds. *Random Sets Theory and Applications*, Springer-Verlag New York (1997).

[307] I. Goodman, R. Mahler, and H. Nguyen. *Mathematics of Data Fusion*, Kluwer Academic Publishers (1997).

[308] H. Grad. Principles of the kinetic theory of gases. In *Flugge's Handbuch des Physik*, Springer-Verlag, vol. XII. pp. 205–294 (1958).

[309] C. Graham and S. Méléard. Stochastic particle approximations for generalized Boltzmann models and convergence estimates *Ann. Probab.*, vol. 25, no. 1, pp. 115–132 (1997).

[310] C. Graham. A statistical physics approach to large networks. Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math., Springer Berlin, vol. 1627, pp.127–147 (1996).

[311] C. Graham and S. Méléard. Dynamic asymptotic results for a generalized star-shaped loss network. *Ann. Appl. Probab.*, vol. 5, no. 3, pp. 666–680 (1995).

[312] C. Graham. Homogenization and propagation of chaos to a nonlinear diffusion with sticky reflection. *Probab. Theory Related Fields*, vol. 101, no. 3, pp. 291–302 (1995).

[313] C. Graham and S. Méléard. Fluctuations for a fully connected loss network with alternate routing. *Stochastic Process. and Appl.*, vol. 53, no. 1, pp. 97–115 (1994).

[314] C. Graham and S. Méléard. Chaos hypothesis for a system interacting through shared resources. *Probab. Theory Related Fields*, vol. 100, no. 2, pp. 157–173 (1994).

[315] C. Graham and S. Méléard. Propagation of chaos for a fully connected loss network with alternate routing. *Stochastic Process. Appl.*, vol. 44, no. 1, pp. 159–180 (1993).

[316] C. Graham. Nonlinear diffusion with jumps. *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 28, no. 3, pp. 393–402 (1992).

[317] C. Graham. McKean-Vlasov Ito-Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stochastic Process. Appl.*, vol. 40, no. 1, pp. 69–82 (1992).

[318] C. Graham. Nonlinear limit for a system of diffusing particles which alternate between two states. *Appl. Math. Optim.*, vol. 22, no. 1, pp. 75–90 (1990).

[319] C. Graham and M. Métivier. System of interacting particles and nonlinear diffusion reflecting in a domain with sticky boundary. *Probab. Theory Related Fields*, vol. 82, no. 2, pp. 225–240 (1989).

[320] C. Graham. The martingale problem with sticky reflection conditions, and a system of particles interacting at the boundary. *Ann. Inst. H. Poincar Probab. Statist.*, vol. 24, no. 1, pp. 45–72 (1988).

[321] I. Grigorescu and M. Kang. Hydrodynamic limit for a Fleming-Viot type system. *Stochastic Process. Appl.*, vol. 110, no. 1, pp. 111–143 (2004).

[322] A. Guyader, N. Hengartner, and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Applied Mathematics & Optimization* (2011).

[323] P. Grassberger. Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1,000,000. *Phys. Rev. E*, pp. 3682–3693 (1997).

[324] P. Grassberger. Go with the winners: A general Monte Carlo strategy. *Computer Physics Communications*, vol. 147, no. 1, pp. 64–70 (2002).

[325] E. J. Green. Continuum and finite player noncooperative models of competition. *Econometrica*, vol. 52, no. 4 pp. 975–993 (1984).

[326] G. Grimmett, and D. Stirzaker. *Probability and Random Processes*, 2nd ed. Oxford University Press, Oxford (1992).

[327] D. Grünbaum. Translating stochastic density-dependent individual behavior with sensory constraints to an Eulerian model of animal swarming. *J. Math. Biology*, vol. 33, pp. 139–161 (1994).

[328] D. Grünbaum and A. Okubo. Modelling social animal aggregations. In *Frontiers of Theoretical Biology* (S. Levin, Ed.), Lectures Notes in Biomathematics, 100, Springer-Verlag, New York, pp. 296–325 (1994).

[329] S. Gueron, S.A. Levin, and D.I. Rubenstein. The dynamics of herds: from individuals to aggregations, *J. Theor. Biol.*, 182, pp. 85–98 (1996).

[330] M. Hairer. Convergence of Markov processes. Lecture notes, Warwick University (2010).

[331] N. M. Haan and S. J. Godsill. Sequential methods for DNA sequencing. Acoustics, Speech, and Signal Processing, 2001. *Proceedings (ICASSP'01). 2001 IEEE International Conference*, Vol. 2. IEEE (2001).

[332] O. H. Hald. Convergence of random methods for a reaction diffusion equation. *SIAM J. Sci. Statist. Comput.*, vol. 2, pp. 85–94 (1981).

[333] P.R. Halmos. Measure theory. v. Nostrand (1950).

[334] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.*, vol. 43, no. 2, pp. 1079–1083 (1971).

[335] T. E. Harris. *The Theory of Branching Processes*, Springer-Verlag, Berlin (1963).

[336] T. E. Harris. Branching processes. *Ann. Math. Statist.*, 19, pp. 474–494 (1948).

[337] J. W. Harris, and S.C Harris. Survival probability for branching Brownian motion with absorption. *Electron. Comm. Probab.*, vol. 12, pp. 89–100 (2007).

[338] T. E. Harris and H. Kahn. Estimation of particle transmission by random sampling. *Natl. Bur. Stand. Appl. Math. Ser.*, vol. 12, pp. 27–30 (1951).

[339] P. H. Haynes and J. Vanneste. What controls the decay rate of passive scalars in smooth random flows? *Phys. Fluids*, vol. 17, 097103 (2005).

[340] K. Heine. Unified framework for sampling/Importance resampling algorithms. *Proceedings of FUSION Conference*, Philadelphia (2005)

[341] A. W. Heemink, M.Verlaan, and A. J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, vol. 129, no. 7, pp. 1718–1728 (2001).

[342] S. Herrmann, and J. Tugaut. Non uniqueness of stationary measures for self-stabilizing diffusions Stochastic Processes & Their Applications Vol. 120, no. 7, pp. 1215–1246 (2010).

[343] J. H. Hetherington. Observations on the statistical iteration of matrices. *Phys. Rev. A*, vol. 30, pp. 2713–2719 (1984).

[344] J. H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor (1975).

[345] C. Houdré. P. Reynaud-Bouret. Exponential inequalities, with constants, for U-statistics of order two. *Stochastic inequalities and applications*, Progr. Probab., 56 Birkhäuser, Basel, pp. 55–69 (2003)

[346] M. Huang, R.P. Malhamé, and P.E. Caines. Large population stochastic dynamical games. Closed loop McKean Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, vol. 6, no. 3, pp. 221–252 (2006).

[347] M. Huang, R.P. Malhamé, and P.E. Caines. Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and Nash equilibrium solutions. *Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii*, December 2003, pp. 98–103 (2003).

[348] M. Huang, R.P. Malhamé, and P.E. Caines. Large-population cost-coupled LQG problems: generalizations to nonuniform individuals. *Proceedings of the 43rd IEEE Conference on Decision and Control, Atlantis, Paradise Island, Bahamas*, December 2004, pp. 3453–3458 (2004).

[349] M. Huang, R.P. Malhamé, P.E. Caines. An invariance principle in large population stochastic dynamic games. *Journal of Systems Science & Complexity*, vol. 20, no. 2, pp. 162–172 (2007).

[350] M. Huang, R.P. Malhamé, and P.E. Caines. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ?-Nash equilibria. *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1560–1571 (2007).

[351] M. Huang, P.E. Caines, and R.P. Malhamé. Social optima in mean field LQG control: centralized and decentralized strategies. *Proceedings of the 47th Annual Allerton Conference, Allerton House, UIUC, Illinois*, September 2009, pp. 322–329 (2009).

[352] M. Huang, P.E. Caines, and R.P. Malhamé, Social certainty equivalence in mean field LQG control: social, Nash and centralized strategies. *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems, Budapest, Hungary*, July 2010, pp. 1525–1532 (2010).

[353] Y. Iba. Population Monte Carlo algorithms. Transactions of the Japanese Society for Artificial Intelligence, vol. 16, no. 2, pp 279–286 (2001).

[354] E. Ikonen, P. Del Moral, and K. Najim. A genealogical decision tree solution to optimal control problems. In *IFAC Workshop on Advanced Fuzzy/Neural Control, Oulu, Finland*, pp. 169–174 (2004).

[355] E. Ikonen, K. Najim, and P. Del Moral. Application of genealogical decision trees for open-loop tracking control. In *Proceedings of the 16th IFAC World Congress, Prague, Czech* (2005).

[356] M. Isard and A. Blake. Condensation conditional probability density propagation for visual tracking. *International Journal of Computer Vision*, vol. 29, pp. 5–28 (1998).

[357] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, New York (1987).

[358] P. Jagers. *Branching Processes with Biological Applications*. John Wiley & Sons, London (1975).

[359] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, vol. 78, 2690 (1997).

[360] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, vol. 56, 5018 (1997).

[361] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York (1970).

[362] M. S. Johannes, N. G. Polson, and J. R. Stroud. Optimal filtering of jump diffusions: extracting latent states from asset prices. *Review of Financial Studies*, 22, issue. 7, pp. 2759–2799 (2009).

[363] A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proceedings of 6th International Workshop on Rare Event Simulation, Bamberg, Germany* (2006).

[364] W. B. Johnson, G. Schechtman, and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.*, vol. 13, no. 1, pp. 234–253 (1985).

[365] B. Jourdain. Diffusion processes associated with nonlinear evolution equations for signed measures. *Methodology and Computing in Applied Probability*, vol. 2, no. 1, pp. 69–91 (2000).

[366] B. Jourdain and F. Malrieu. Propagation of chaos and Poincar inequalities for a system of particles interacting through their cdf. *Annals of Applied Probability*, vol. 18, no. 5, 1706–1736 (2008).

[367] M. Kac. On distributions of certain Wiener functionals, *Trans. American Math. Soc.*, vol. 61, no. 1, pp. 1–13, (1949).

[368] M. Kac. Foundations of kinetic theory. *Proceedings 3rd Berkeley Sympos. Math. Stat. and Proba.*, vol. 3, pp. 171-197 (1956).

[369] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York (1997).

[370] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering, Series D.*, vol. 82, pp. 25–45 (1960).

[371] I. Karatzas and S.E. Shreve. *Methods of Mathematical Finance*. Springer (1998).

[372] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer (2004).

[373] S. Karlin. *Mathematical Methods and Theory in Games, Programming and Economics*, in two volumes, Addison-Wesley, London (1959).

[374] S. Karlin. Stochastic models and optimal policy for selling an asset. In *Studies in Applied Probability and Management Science*. K. J. Arrow, S. Karlin, and H. Scarf, Eds., Stanford University Press, 148-158 (1962).

[375] H. Kesten. Branching Brownian motion with absorption. *Stochastic Process. and Appl.*, vol. 37, pp. 9–47 (1978).

[376] J. F. C. Kingman. The first birth problem for age dependent branching processes. *Ann. Probab.*, vol. 3, pp. 790–801 (1975).

[377] V. N. Kolokoltsov. *Nonlinear Markov Processes and Kinetic Equations*. Cambridge Univ. Press (2010).

[378] V. N. Kolokoltsov, J. Li, W. Yang. Mean field games and nonlinear Markov processes. arXiv:1112.3744v2 (2012).

[379] V. N. Kolokoltsov, O. A. Malafeyev. *Analysis of Many Agent Systems of Competition and Cooperation. Game Theory for All* (in Russian). St. Petersburg University Press (2007).

[380] V. N. Kolokoltsov, O.A. Malafeyev. *Understanding Game Theory. Introduction to the Analysis of Many Agent Systems of Competition and Cooperation*. World Scientific (2010).

[381] H. R. Künsch. State-space and hidden Markov models. In *Complex Stochastic Systems*. O. E. Barndorff-Nielsen, D. R. Cox, and C. Kluppelberg, Eds., CRC Press, pp. 109–173 (2001).

[382] J.G. Kemeny and J. L. Snell. *Finite Markov Chains*. Springer-Verlag, New York (1976).

[383] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comp. Graph. Statist.*, vol. 5, pp. 1–25 (1996).

[384] P. E. Kloeden, E. Platen, and N. Hofmann. Extrapolation methods for the weak approximation of Itô diffusions. *SIAM Journal on Numerical Analysis*, vol. 32, no. 5, pp. 1519–1534 (1995).

[385] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, vol. 23, Springer (2011).

[386] P. E. Kloeden and E. Platen. A survey of numerical methods for stochastic differential equations. *Stochastic Hydrology and Hydraulics*, vol. 3, no. 5, pp. 155-178 (1989).

[387] E. Platen and K. Kubilius. Rate of weak convergence of the Euler approximation for diffusion processes with jumps. School of Finance and Economics, University of Technology, Sydney (2001).

[388] A. Kohatsu-Higa and M. Montero. *Malliavin Calculus in Finance*. Handbook of Computational and Numerical Methods in Finance, Birkhäuser, pp. 111–174 (2004).

[389] A. Kohatsu-Higa and S. Ogawa. Weak rate of convergence for an Euler scheme of Nonlinear SDE's. In *Monte Carlo Methods and Its Applications*, vol. 3, pp. 327–345 (1997).

[390] V. N. Kolokoltsov and V. P. Maslov. *Idempotent Analysis and Its Applications*, vol. 401 of *Mathematics and its Applications* with an appendix by P. Del Moral on Maslov optimization theory. Kluwer Academic Publishers Group, Dordrecht (1997).

[391] K. Kremer and K. Binder. Monte Carlo simulation of lattice models for macromolecules. *Computer Physics Reports*, vol. 7, no. 6, pp. 259–310 (1988).

[392] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*, Wiley Series Probability & Statistics (1986).

[393] A. Lachapelle, J. Salomon, and G. Turinici. Computation of mean field equilibria in economics. *Mathematical Models and Methods in Applied Sciences*, vol. 20, no. 4, pp. 567–588 (2010).

[394] T. Laffargue, K. D. Nguyen Thu Lam, J. Kurchan, and J. Tailleur. Large deviations of Lyapunov exponents. ArXiv 1302.6254 (2013)

[395] L. D. Landau, E. M. Lifshitz. Course of theoretical physics. *Physical kinetics*, E. M. Lifshitz and L. P. Pitaevskii eds. vol. 10, Oxford, Pergamon (1981).

[396] O.E. Lanford and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Physics*, vol. 13, pp. 194–215 (1969).

[397] J. M. Laskry and P. L. Lions. Contrôle stochastique avec informations partielles et applications à la Finance. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, vol. 328, issue 11, pp. 1003–1010 (1999).

[398] J. M. Lasry and P.L. Lions. Mean field games. *Japanese J. Math*, vol. 2, no. 1, pp. 229–260 (2007).

[399] J. M. Lasry and P.L. Lions. Jeux à champ moyen. I. Le cas stationnaire. *C.R. Math. Acad. Sci. Paris*. 343(10), pp. 619-625 (2006).

[400] J. M. Lasry and P.L. Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C.R. Math. Acad. Sci. Paris.* 343(10), pp. 679–684 (2006).

[401] J. M. Lasry and P.L. Lions. Mean field games. *Cahiers de la Chaire Finance et Développement Durable* (2) (2007).

[402] D. Lefèvre. An introduction to utility maximization with partial observation. *Finance*, vol. 23 (2002).

[403] F. Le Gland and N. Oudjane, A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Processes and Their Applications*, vol. 106, no. 2, pp. 279–316 (2003).

[404] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Annals Applied Probability*, vol. 14, no. 1, 144–187 (2004).

[405] F. Le Gland, V. Monbet, and V. D. Tran. Large sample asymptotics for the ensemble Kalman filter. In *Handbook on Nonlinear Filtering*, D. Crisan and B. Rozovskii, Eds. Oxford University Press, Oxford, pp. 598–631 (2011).

[406] V. Tran, V. Monbet and F. Le Gland. Filtre de Kalman d'ensemble et filtres particulaires pour le modèle de Lorenz. Actes de la Manifestation des Jeunes Chercheurs STIC (MajecSTIC'06), Lorient, November 22-24 (2006).

[407] M. Lei and C. Baehr. Unscented and ensemble transform-based variational filter. Accepted to publication in *Physica D : Nonlinear Phenomena* (2012).

[408] M. Ledoux. The concentration of measure phenomenon. *AMS Monographs, Providence* (2001).

[409] T. Lelièvre, M. Rousset, and G. Stoltz. Computation of free energy differences through nonequilibrium stochastic dynamics: the reaction coordinate case. *J. Comp. Phys.*, vol. 222, no. 2, pp. 624–643 (2007).

[410] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations: A mathematical perspective.* 472 pp., Imperial College Press (2010).

[411] C. Léonard. Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coéfficients non bornés. *Ann. Inst. Henri Poincaré*, vol. 22, no. 2, pp. 237–262 (1986).

[412] T. Li, J.F. Zhang. Asymptotically optimal decentralized control for large population stochastic multiagent systems. *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1643–1660 (2008).

[413] L. Li, J. Bect, and E. Vazquez. Bayesian subset simulation: a kriging-based subset simulation algorithm for the estimation of small probabilities of failure. arXiv:1207.1963 v1 (2012).

[414] C. Liang, G. Cheng, D. L. Wixon. T. C. Balser. An absorbing Markov chain approach to understanding the microbial role in soil carbone stabilization. *Biogeochemistry*, vol. 106, pp. 303–309 (2011).

[415] A. M. L. Liekens. Evolution of Finite Populations in Dynamic Environments. Ph.D. thesis, Technische Universiteit Eindhoven (2005).

[416] M. Liebscher, S. Pannier, J. Sickert, and W. Graf. Efficiency improvement of stochastic simulations by means of subset sampling, *In Proceedings of the 5th German LS-DYNA Forum 2006. DYNAmore GmbH*, Ulm (2006).

[417] A. E. B. Lim. Mean variance hedging when there are jumps. *SIAM J. Control Optim.*, vol. 44, no. 5, pp. 1893–1922 (2005).

[418] P.L. Lions. Théorie des jeux à champ moyen et applications. Cours au Collège de France (2007-2008).

[419] P.L. Lions. *Mathematical Topics in Fluid Mechanics.* Oxford Science Publications. Clarendon Press. Oxford, vol. 1 (1996), vol 2 (1998).

[420] J.S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer Verlag, New York 2001.

[421] G. Liu and L. J. Hong. Revisit of stochastic mesh method for pricing American options. *Operations Research Letters*, vol. 37, no. 6, pp. 411–414 (2009).

[422] J.S. Liu and J.Z. Zhang. A new sequential importance sampling method and its applications to the 2-dimensional hydrophobichydrophilic model. *J. Chem. Phys.*, vol. 117, no. 7, pp. 3492–3498 (2002).

[423] J.S. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *J. Am. Stat. Assoc.*, vol. 93, pp. 1032–1044 (1998).

[424] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, vol. 14, pp. 113–147 (2001).

[425] H. F. Lopes, N. G. Polson, and C. M. Carvalho. Bayesian statistics with a smile: A resampling-sampling perspective. *Brazilian Journal of Probability and Statistics*, vol. 26, no. 4, pp. 358–371 (2012).

[426] S. McGinnity. Multiple model bootstrap filter for maneuvering target tracking. *IEEE Transactions on Aerospace and Electronic Systems,* vol. 36, no. 3, pp. 1006–1012 (2000).

[427] R. Mahler. A theoretical foundation for the Stein-Winter Probability Hypothesis Density (PHD) multi-target tracking approach, *Proc. MSS Nat'l Symp. on Sensor and Data Fusion,* vol. I, San Antonio TX (2000).

[428] R. Mahler. Multi-target Bayes filtering via first-order multi-target moments, *IEEE Trans. Aerospace & Electronic Systems*, vol. 39, no. 4, pp. 1152–1178 (2003).

[429] R. Mahler. Global integrated data fusion, *Proc. 7th Nat. Symp. on Sensor Fusion*, vol. 1, Sandia National Laboratories, Albuquerque, ERIM Ann Arbor, MI, pp. 187–199 (1994).

[430] R. Mahler. Random-set approach to data fusion, *SPIE*, vol. 2234 (1994).

[431] P. Major. Estimation of multiple random integrals and U-statistics. *Moscow Mathematical Journals*, vol. 10 no. 4, pp. 747–763 (2010).

[432] P. Major. Tail behaviour of multiple random integrals and U-statistics. *Probability Reviews Series*, vol. 2 pp. 448–505 (2005).

[433] P. Major. An estimate about multiple stochastic integrals with respect to a normalized empirical measure. *Studia Scientarum Mathematicarum Hungarica*, vol. 42, no. 3, pp. 295-341 (2005).

[434] P. Major. A multivariate generalization of Hoeffding's inequality. *Electronic Communication in Probability*, vol. 2, pp. 220–229 (2006).

[435] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes *Annals of Applied Probability*, vol. 13, no. 2, pp. 540–560 (2003).

[436] F. Malrieu. Inégalités de Sobolev logarithmiques pour des problèmes d'évolutions non linéaires. Ph.D. dissertation, Hal-00001287. *Université Paul Sabatier* (2001).

[437] S. Martinez, S. San Martin, and D. Villemonais. Existence and uniqueness of quasi- stationary distributions and fast return from infinity. Submitted to *Journal of Applied Probability*, 15 p. (2012).

[438] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, vol. 101, no. 2, pp. 185–232 (2002).

[439] F. A. Matsen and J. Wakeley. Convergence to the island-model coalescent process in populations with restricted migration. *Genetics*, vol. 172, pp. 701–708 (2006).

[440] J. C. Maxwell. On the dynamical theory of gases. *Philos. Trans. Roy. Soc. London Ser. A*, vol. 157, pp. 49–88 (1867).

[441] J. C. Maxwell. On stresses in rarified gased arising from inequalities of temperatures. *Philos. Trans. Roy. Soc. London Ser. A*, vol. 170, pp. 231–256 (1879).

[442] J. C. Maxwell. *The Scientific Letters and Papers of James Clerk Maxwell.* vol. 2., pp. 1862-1873, Cambridge University Press (1995).

[443] J. Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, Cambridge (1982).

[444] L. Mazliack. Approximation of a partially observable stochastic control problem. *Markov Processes and Related Fields*, vol. 5, pp. 477–486 (1999).

[445] H. P. McKean Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. USA*. vol. 56, pp. 1907–1911 (1966).

[446] S. Méléard. Probabilistic interpretation and approximations of some Boltzmann equations. In Stochastic models. *Soc. Mat. Mexicana, Mexico*, pp. 1–64 (1998).

[447] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. Probabilistic models for nonlinear partial differential equations. Lecture Notes in Mathematics, vol. 1627-1996, pp. 42–95 (1996).

[448] S. Méléard and S. Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. *Stochastic Process. Appl.*, vol. 26, pp. 317–332, (1987).

[449] S. Méléard. and D. Villemonais. Quasi-stationary distributions and population processes. Submitted to *Probability Surveys*, 65 p. (2011).

[450] V. Melik-Alaverdian and M.P. Nightingale. Quantum Monte Carlo methods in statistical mechanics, *Internat. J. of Modern Phys. C*, vol. 10, pp. 1409–1418 (1999).

[451] M. Métivier. Quelques problèmes liés aux systèmes infinis de particules et leurs limites. Lecture Notes in Math., vol. 1204, pp. 426–446 (1984).

[452] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341 (1949).

[453] N. Metropolis. The Beginning of the Monte Carlo method. *Los Alamos Science*, no. 15, pp. 125–130 (1987).

[454] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092 (1953).

[455] S. P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London. Available at: probability.ca/MT (1993).

[456] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press (2007).

[457] S. Mischler, C. Mouhot. Kac's program in kinetic theory. arXiv:1107.3251 (2011), to appear in *Inventiones mathematicae* (2013).

[458] S. Mischler, C. Mouhot, B. Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. arXiv:1101.4727 (2011).

[459] D. Morale, V. Capasso, and K. Oeslschläger. An interacting particle system modeling aggregation behavior: from individual to populations. *J. Math. Bio.*, vol. 50, pp. 49–66 (2005).

[460] J. E. Moyal. The general theory of stochastic population processes. *Acta Mathematica*, vol. 108, pp. 1–31 (1962).

[461] K. P. Murphy, A. Doucet, N. De Freitas, and S. Russell. Rao-Blackwellised particle filtering for dynamic Bayesian networks. in *Proc. Uncertainty in Artificial Intelligence* (2000).

[462] T. Nagai and M. Mimura. Asymptotic behaviour for a nonlinear degenerate diffusion equation in population dynamics. *SIAM J. Appl. Math.*, 43, pp. 449–464 (1983).

[463] T. Nagai and M. Mimura. Some nonlinear degenerate diffusion equations related to population dynamics, *J. Math. Soc. Japan*, 35, pp. 539–561 (1983).

[464] K. Najim, E. Ikonen, and P. Del Moral. Open-loop regulation and tracking control based on a genealogical decision tree. *Neural Computing & Applications*, vol. 15, no. 3–4, pp. 339–349 (2006).

[465] J. Neveu. Multiplicative martingales for spatial branching processes. In *Seminar on Stochastic Processes, 1987*. E. Cinlar, K. L. Chung and R. K. Getoor, eds. *Prog. Probab. Statist.* vol.15, pp. 223–241. Birkhäuser, Boston (1988).

[466] M. Nourian, P.E. Caines, and R.P. Malhamé. Mean field analysis of controlled Cucker-Smale type flocking: linear analysis and perturbation equations. *Proceedings of the 18th IFAC World Congress, Milan*, August 2011, pp. 4471–4476 (2011).

[467] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, vol. 12, pp. 458–479 (1984).

[468] K. Oelschläger. On the derivation of reaction-diffusion equations as limit of dynamics of systems of moderately interacting stochastic processes. *Prob. Th. Rel. Fields*, vol. 82, pp. 565–586 (1989).

[469] K. Oelschläger. Large systems of interacting particles and porous medium equation, *J. Differential Equations*, vol. 88, pp. 294–346 (1990).

[470] B. K. Øksendal and A. Sulem. *Applied Stochastic Control of Jump Diffusions*. Springer, Berlin (2007).

[471] B. K. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. Springer, Berlin (2003).

[472] A. Okubo. Dynamical aspects of animal grouping: swarms, school, flocks and herds. *Adv. Bio-Phys.*, vol. 22, pp. 1–94 (1986).

[473] T.H. Otway. The Dirichlet problem for elliptic-hypebolic equations of Keldysh type. Lecture Notes in mathematics, Springer (2012).

[474] J. Olsson and T. Rydén. Rao-Blackwellisation of particle Markov chain Monte Carlo methods using forward filtering backward sampling. *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4606–4619 (2011).

[475] A. Osekowski. Two inequalities for the first moments of a martingale, its square function and its maximal function *Bull. Polish Acad. Sci. Math.*, vol. 53, pp. 441–449 (2005).

[476] M. Pace, P. Del Moral, and F. Caron. Comparison of Implementations of Gaussian mixture PHD filters, *Proceedings of the 13th International Conference on Information Fusion* 26-29 July 2010 EICC Edinburgh, UK (2010).

[477] M. Pace. Comparison of PHD based filters for the tracking of 3D aerial and naval scenarios. 2010 IEEE International Radar Conference Washington, DC, on May 10–14 (2010).

[478] M. Pace and P. Del Moral. Mean-field PHD filters based on generalized Feynman-Kac flow. *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2 (2013).

[479] G. Pagès and B. Wilbertz. Optimal quantization methods for pricing American style options. *Numerical Methods in Finance* (2011).

[480] G. Pagès and H. Pham. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, vol. 11, pp. 893–932 (2005).

[481] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of Computational and Numerical Methods in Finance.* S. T. Rachev, Ed. Birkhäuser, Boston, pp. 253–297 (2004).

[482] G. Pagès. A space vector quantization method for numerical integration. *Journal of Comput. Appl. Math.*, 89, pp. 1–38 (1997).

[483] G. Pagès and J. Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Appl.*, vol. 11, no. 4, pp. 407–446 (2005).

[484] N. Papageorgiou, B. Rémillard, and A. Hocquard. Replicating the properties of hedge fund returns. *Journal of Alternative Investments*, vol. 11, pp. 8–38 (2008).

[485] A. Papoulis. *Brownian Movement and Markov Processes. Probability, Random Variables, and Stochastic Processes*, 2nd ed. McGraw-Hill, New York, pp. 515–553 (1984).

[486] E. Pardoux, D. Talay. Discretization and simulation of stochastic differential equations. *Acta Applic. Math.*, vol. 3, no. 1, pp. 22–47 (1985).

[487] G. Peskir, and A. Shiryaev. *Optimal Stopping and Free-Boundary Problems.* Birkhäuser, Basel (2006).

[488] G. Pflug. *Stochastic Optimization.* Kluwer (1997).

[489] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods and Applications*, vol. 11, no. 1, pp. 57–81 (2005).

[490] H. Pham. Mean-variance hedging under partial observation. *International Journal of Theoretical and Applied Finance*, vol. 4, pp. 263–284 (2001).

[491] H. Pham, and M. C. Quenez. Optimal portfolio in partially observed stochastic volatility models. *Annals of Applied Probability*, vol. 11, pp. 210–238 (2001).

[492] H. Pham, M. Corsi, and W. Runggaldier. Numerical approximation by quantization of control problems in finance under partial observations. In *Handbook of Numerical Analysis*, vol. 15, pp. 325–360 (2009).

[493] P. Pérez, A. Blake, and M. Gangnet. Jetstream: probabilistic contour extraction with particles. In *Proceedings, IEEE International Conference on Computer Vision*, pp. 524–531 (2001).

[494] B. Perthame. *Transport Equations in Biology. Frontiers in Mathematics.* Birkhaüser Verlag (2007).

[495] V. M. Pillai and M. P. Chandrasekharan. An absorbing Markov chain model for production systems with rework and scrapping. *Computers and Industrial Engineering*, vol. 55, no. 3, pp. 695–706 (2008).

[496] V. Plachouras, I. Ounis, and G. Amati. The static absorbing model for hyperlink analysis on the web. *Journal of Web Engineering*, vol. 4, no. 2, pp. 165–186 (2005).

[497] D. Pollard. *Empirical processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics. Publication of the Institute of Mathematical Statistics, Hayward, CA, vol. 2 (1990).

[498] R. B. Potts. Some Generalized Order-Disorder Transformations. Mathematical Proceedings, vol. 48, no. 1, pp. 106–109 (1952).

[499] S.T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York (1991).

[500] T. Rainsford and A. Bender. Markov approach to percolation theory based propagation in random media. *IEEE Transactions on Antennas and Propagation.* DOI:10.1109/TAP.2008.922626 (2008) vol 56, no. 3, pp. 1402–1412, May (2008).

[501] B. R. Rambharat and A. E. Brockwell. Sequential Monte Carlo pricing of American-style options under stochastic volatility models. *The Annals of Applied Statistics*, vol. 4, no. 1, pp. 222–265 (2010).

[502] C. R. Reeves and J. E. Rowe. *Genetic Algotihms : Principles and Perspectives: A Guide to GA Theory.* Kluwer Academic Publishers (2003).

[503] B. Rémillard, A. Hocquard, H. Langlois, and N. Papageorgiou. Optimal hedging of American options in discrete time. *Numerical Methods in Finance* (2011).

[504] B. Rémillard and S. Rubenthaler. Optimal hedging in discrete and continuous time. *Technical Report G-2009-77*, Cahiers du Gerad (2009).

[505] S. Rémy, O. Pannekoucke, T. Bergot, and C. Baehr. Adaptation of a particle filtering method for data assimilation in a 1D numerical model used for fog forecasting. *Quarterly Journal of Royal Meteorological Society.* Vol 138, no. 663, pp. 536–551, January (2012).

[506] D. Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York (1991).

[507] C. Reyl, T. M. Antonsen, and E. Ott. Vorticity generation by instabilities in chaotic fluid flows, *Physica D*, vol. 111, pp. 202–226 (1998).

[508] G. Ridgeway and D. Madigan. A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery*, vol. 7, no. 3, pp. 301–319 (2003).

[509] B. D. Ripley. *Stochastic Simulation.* Wiley & Sons (1987).

[510] E. Rio. Local invariance principles and their applications to density estimation. *Probability and Related Fields*, vol. 98, pp. 21–45 (1994).

[511] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*, 2nd ed. Springer-Verlag (2004).

[512] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, vol. 2, pp. 341–363 (1996).

[513] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, vol. 1, pp. 95–110 (1996).

[514] R. Rogerson, R. Shimer, and R. Wright. Search-theoretic models of the labor market: a survey. *Journal of Economic Literature*, 43, pp. 959–988 (2005).

[515] V. Rossi and J. P. Vila. Nonlinear filtering in discrete time: a particle convolution approach. *Ann. I.SU.P.*, vol. 50, no. 3, pp. 71–102 (2006).

[516] M. Rousset. On the control of an interacting particle approximation of Schroedinger ground states, *SIAM J. Math. Anal.*, vol. 38, no. 3, pp. 824–844, (2006).

[517] M. Rousset and G. Stoltz. Equilibrium sampling from nonequilibrium dynamics. *J. Stat. Phys.*, vol. 123, no. 6, pp. 1251–1272 (2006).

[518] R. Rubinstein. Entropy and cloning methods for combinatorial optimization, sampling and counting using the Gibbs sampler. *Information Theory and Statistical Learning*, pp. 385–434 (2009).

[519] R. Rubinstein. The Gibbs cloner for combinatorial optimization, counting and sampling. *Methodology and Computing in Applied Probability*, vol. 11, no. 4, pp. 491–549 (2009).

[520] R. Rubinstein. Randomized algorithms with splitting: why the classic randomized algorithms do not work and how to make them work. *Methodology and Computing in Applied Probability*, vol. 12, no. 1, pp. 1–50 (2010).

[521] R. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley (1981).

[522] R. Rubinstein. *Monte Carlo Optimization Simulation and Sensitivity of Queueing Networks*. Wiley (1986).

[523] R. Rubinstein and D. P. Kroese. *The Cross-Entropy Method. A Unified Approach to Combinatorial Optimization, Simulation and Machine Learning*. Springer-Verlag (2005).

[524] R. Rubinstein, B. Melamed. Classical and Modern Simulation. Wiley (1998).

[525] W. Runggaldier and L. Stettner. *Approximations of Discrete Time Partially Observed Control Problems*. Applied Mathematics Monographs CNR, Giardini Editori, Pisa (1994).

[526] A.Scemama, M. Caffarel, E. Oseret, and W. Jalby. Quantum Monte Carlo for large chemical systems: Implementing efficient strategies for petascale platforms and beyond. Submitted to J.Comp.Chem preprint Uni. Paul Sabatier.

[527] A. Scemama, M. Caffarel, E. Oseret and W. Jalby. QMC=Chem: a quantum Monte Carlo program for large-scale simulations in chemistry at the petascale level and beyond. VECPAR 2012 in press.

[528] S. F. Schmidt. The Kalman filter: recognition and development for aerospace applications. *Journal of Guidance and Control*, vol. 4, no. 1, pp. 4–8 (1981).

[529] E. Seneta. *Nonnegative Matrices and Markov Chains*, 2nd rev. ed., XVI, p. 288, Springer Series in Statistics (1981).

[530] S.S. Singh, B.N. Vo, A. Baddeley and S. Zuyev. Filters for Spatial Point Processes. *SIAM J. Control and Optimization*, vol. 47, no. 4, pp. 2275–2295 (2009).

[531] C. Schäfer and N. Chopin. Adaptive Monte Carlo on binary sampling spaces. *Statistics and Computing*. doi:10.1007/s11222-011-9299-z (2012).

[532] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. Physical Review, vol. 28, no. 6, pp. 1049–1070 (1926).

[533] M. Schweizer. Variance-optimal hedging in discrete time. *Math. Oper. Res.*, vol. 20, no. 1, pp. 1–32 (1995).

[534] O. Schutze, C. A. Coello Coello, A.A. Tantar, E. Tantar, P. Bouvry, P. Del Moral, and P. Legrand. *Evolve - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation II*. Advances in Intelligent Systems and Computing, Volume 175, Springer (2012).

[535] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass, *Physics Review Letters*, vol. 35, no. 26, pp. 1792–1796 (1975).

[536] Z. Shi. Branching random walks. École d'été de Probabilités de Saint Flour (2012).

[537] T. Shiga and H. Tanaka. Central limit theorems for a system of Markovian particles with mean field interactions. *Z. Wahrscheinlichkeitstheorie, verw. Gebiete*, vol. 69, pp. 439–459 (1985).

[538] A. N. Shiryaev. On optimal methods in quickest detection problems. *Theory Prob. and Appl.*, vol. 8, pp. 22–46 (1963).

[539] A. N. Shiryaev. *Probability*. Graduate Texts in Mathematics, 2nd ed., vol. 95, Springer (1996).

[540] O. Skare, E. Bolviken, and L. Holden. Improved sampling-importance resampling and reduced bias importance sampling. *Scandinavian Journal of Statistics*, vol. 30, no. 4, pp. 719–737 (2003).

[541] Z. Skolicki and K. De Jong. The influence of migration sizes and intervals on island models. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2005).* ACM Press (2005).

[542] A. F. M. Smith and G. O. Roberts. Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B.*, vol. 55, no. 1, pp. 3–23 (1993).

[543] A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, vol. 46, no. 2, pp. 84–88 (1992).

[544] P. J. Smith, M. Shafi, and H. Gao. Quick simulation: A review of importance sampling techniques in communication systems. *IEEE J. Select. Areas Commun.*, vol. 15, pp. 597–613 (1997).

[545] L. J. Snell. Applications of martingale system theorems. *Trans. Amer. Math. Soc.*, vol. 73, pp. 293–312 (1952).

[546] R. Srinivasan. *Importance Sampling - Applications in Communications and Detection*, Springer-Verlag, Berlin (2002).

[547] H. E. Stanley. *Mean Field Theory of Magnetic Phase Transitions. Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press (1971).

[548] N. Starr. How to win a war if you must: optimal stopping based on success runs. *Ann. Math. Statist.*, vol. 43, pp. 1884–1893, (1972).

[549] N. Starr. Optimal and adaptive stopping based on capture times. *J. Appl. Prob.*, vol. 11, pp. 294–301 (1974).

[550] N. Starr and M. Woodroofe. Gone fishin: Optimal stopping based on catch times. *Univ. Mich. Tech. Report*, no. 33, Dept. of Statistics (1974).

[551] N. Starr, R. Wardrop, and M. Woodroofe. Estimating a mean from delayed observations. *Z. fur Wahr.*, vol. 35, pp. 103–113 (1976).

[552] D. Steinsaltz and S. N. Evans. Markov mortality models: implications of quasi-stationarity and varying initial distributions. *Theor. Pop. Biol.*, vol. 65, pp. 319–337 (2004).

[553] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed., Wiley (1995).

[554] R. L. Stratonovich. Conditional Markov processes. *Theory of probability and its applications*, vol. 5, pp. 156–178 (1960).

[555] R. L. Stratonovich. Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise. *Radiofizika*, vol. 2, no. 6, pp. 892–901 (1959).

[556] R. L. Stratonovich. On the theory of optimal nonlinear filtering of random functions. *Theory of Probability and Its Applications*, 4, pp. 223–225 (1959).

[557] R. L. Stratonovich. Application of the Markov processes theory to optimal filtering. *Radio Engineering and Electronic Physics*, vol. 5, no. 11, pp. 1–19 (1960).

[558] F. Suzat, C. Baehr, and A. Dabas. A fast atmospheric turbulent parameters estimation using particle filtering. Application to LIDAR observations. *J. Phys. Conf. Ser.*, vol. 318-072019 doi:10.1088/1742-6596/318/7/072019 (2011).

[559] A. S. Sznitman. Topics in propagation of chaos, course given at the Saint-Flour Probability Summer School, 1989. Lecture Notes in Math., 1464, Springer, Berlin, pp. 164–251 (1991).

[560] W. J. Stewart. *Introduction to the Numerical Solution of Markov Chains.* Princeton University Press, Princeton, NJ (1995).

[561] J. Tailleur, S. Tanese-Nicola, J. Kurchan. Kramers equations an supersymmetry. *Journal Stat. Phys.*, vol. 122, pp. 557–595 (2006).

[562] J. Tailleur, J. Kurchan. Probing rare physical trajectories with Lyapunov weighted dynamics. *Nature Physics*, vol.3 pp. 203–207 (2007).

[563] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, no. 81, pp. 73–205 (1995).

[564] M. Talagrand. A new look at independence. *Ann. Probab.*, vol. 24, pp. 1–34 (1996).

[565] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, vol. 125, pp. 505–563 (1996).

[566] H. Tanaka. Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete*, vol. 46, no. 1, pp. 67–105 (1979).

[567] S. Tanese-Nicola, and J. Kurchan. Metastable states, transitions, basins and borders at finite temperature. *Journal Stat. Phys.*, vol. 116, no. 5–6, pp. 1202–1245 (2004).

[568] E. Tantar, A. Tantar, P. Bouvry, P. Del Moral, C. A. Coello Coello, P. Legrand, and O. Schutze. *Evolve: A bridge between Probability, Set Oriented Numerics and Evolutionary Computation I.* Springer Series: Studies in Computational Intelligence, vol. 447, no. XII, Springer (2013).

[569] H. Tembine, Q. Zhu, and T. Basar. Risk-sensitive mean-field stochastic differential games. *Proceedings of the 18th IFAC World Congress*, Milan, August 2011, pp. 3222–3227 (2011).

[570] J. L. Thorne, H. Kishino, and I.S. Painter. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, vol. 15, no. 12, pp. 1647–1657 (1998).

[571] S. T. Tokdar and R. E. Kass. Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 54–60 (2010).

[572] M. D. Towler. Quantum Monte Carlo and the CASINO program : highly accurate total energy calculations for finite and periodic systems. Psi-k Newsletter "Scientific Highlight of the Month". December (2003).

[573] V. D. Tran. Assimilation de données: les propriétés asymptotiques du filtre de Kalman d'ensemble. *Université de Bretagne Sud*, June 29 (2009).

[574] Y. K. Tsang, E. Ott, T. M. Antonsen, and P. N. Guzdar. Intermittency in two-dimensional turbulence with drag. *Phys. Rev. E*, vol. 71, 066313 (2005).

[575] J. N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, vol. 12, no. 4 (special issue on computational finance), pp. 694–703 (2001).

[576] G. E. Uhlenbeck, G. W. Ford. *Lectures in Statistical Mechanics*. American Mathematical Society, Providence, RI (1963).

[577] R. Van Handel. Uniform time average consistency of Monte Carlo particle filters. *Stoch. Proc. Appl.*, 119, pp. 3835–3861 (2009).

[578] J. Vanneste. Estimating generalized Lyapunov exponents for products of random matrices. *Phys. Rev. E*, vol. 81, 036701 (2010).

[579] F. Viens. Portfolio optimization under partially observed stochastic volatility. *Proceedings of the 8th International Conference on Advances in Communication and Control: Telecommunications / Signal Processing*, pp. 3–12, (2001).

[580] J. P. Vila and V. Rossi. Nonlinear filtering in discret time: a particle convolution approach. *Biostatistic group of Montpellier, Technical Report 04-03* (available at http://vrossi.free.fr/recherche.html) (2004).

[581] A. N. Van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer Series in Statistics. Springer, New York (1996).

[582] P. Varaiya. N-players stochastic differential games. *SIAM J. Control Optim.*, vol. 14, pp. 538–545 (1976).

[583] D. Villemonais. Interacting particle systems and Yaglom limit approximation of diffusions with unbounded drift. *Electronic Journal of Probability*, vol. 16 (2011).

[584] D. Villemonais. Approximation of quasi-stationary distributions for 1-dimensional killed diffusions with unbounded drifts. arXiv:0905.3636 (2009).

[585] D. Villemonais. Interacting particle processes and approximation of Markov processes conditioned to not be killed. In revision for *ESAIM: Probability and Statistics*, p. 26 (2011).

[586] B. T. Vo, B. N. Vo, and A. Cantoni. The cardinalized probability hypothesis density filter for linear Gaussian multi-target models, *Proc. 40th Conf. on Info. Sciences & Systems*, Princeton, (2006).

[587] B. T. Vo, B. N. Vo, and A. Cantoni. Analytic implementations of the Cardinalized Probability Hypothesis Density Filter. *IEEE Trans. Signal Processing*, vol. 55, no. 7, part 2, pp. 3553–3567 (2007).

[588] B. N. Vo and W. K. Ma. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Processing, IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4091–4104 (2006).

[589] B. T. Vo. Random finite sets in multi-object filtering, Ph.D. thesis, University of Western Australia (2008).

[590] K. Wolny. Geometric ergodicity of heterogeneously scaled Metropolis-adjusted Langevin algorithms (MALA). Poster Seventh Workshop on Bayesian Inference in Stochastic Processes (BISP 2011), Madrid, Spain, 1-3.09.2011 (2011).

[591] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory.* MIT Press (1999).

[592] A. Wald. *Sequential Analysis*, John Wiley & Sons, New York (1947).

[593] A. Wald. *Statistical Decision Functions*, John Wiley & Sons, New York (1950).

[594] D. Whitley, S. Rana, and R. B. Heckendorn. The island model genetic algorithm: on separability, population size and convergence. *Journal of Computing and Information Technology,* vol. 7, no. 1, pp. 33–47 (1999).

[595] A. H. Wright and M. D. Vose. Stability of vertex fixed points and applications. In *Foundations of Genetic Algorithms 3*, Morgan Kaufman Publishers, vol. 3 (1995).

[596] A. M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, vol. 56, pp. 795–798 (1947).

[597] T. Yang., P.G. Metha, and S.P. Meyn. A mean-field control-oriented approach to particle filtering. *Proceedings of the American Control Conference (ACC)*, June 2011, pp. 2037–2043 (2011).

[598] L. Zhang and S. Dai. Application of Markov model to environment fate of phenanthrene in Lanzhou reach of Yellow river. *Chemosphere*, vol. 67, pp. 1296–1299 (2007).